Queen’s Economics Department Working Paper No. 1363

The cointegrated vector autoregressive model with general deterministic terms

Søren Johansen
University of Copenhagen and CREATES

Morten Ørregaard Nielsen
Queen’s University and CREATES

Department of Economics
Queen’s University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

7-2016

The cointegrated vector autoregressive model with general deterministic terms

Søren Johansen*
University of Copenhagen
and CREATES

Morten Ørregaard Nielsen†
Queen's University
and CREATES

June 4, 2017

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 models in many areas of empirical economics, as well as other disciplines. However, the formulation and modeling of deterministic terms in the CVAR model have until now been analyzed only 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 simulations, and forecasting are much more sensitive to the specification of the deterministic than the stochastic components of the VAR model."

*Corresponding author. Address: 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

†Email: mon@econ.queensu.ca

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 innovation formulations. In the additive formulation, the deterministic terms are added to the process, and in the innovation formulation they are added to the dynamic 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. We let $\pi = (\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1})$ denote the parameters in $\Pi(L)$ and let $\lambda = (\pi, \gamma)$ with true value $\lambda_0 = (\pi_0, \gamma_0)$. Then λ consists of freely varying parameters, where α, β are $p \times r$ for some $r < p$. Throughout, we will use the Gaussian likelihood function to derive (quasi-) maximum likelihood estimators (MLEs), but as usual, asymptotic properties will not require normality. We also fix $\Omega = \Omega_0$ at the true value, which is without loss of generality in the asymptotic analysis of the remaining parameters because inference on Ω is asymptotically independent of inference on λ.

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 identify $\beta'\gamma_1$.

When analyzing properties of the process, the following $I(1)$ conditions are important, see Johansen (1996, Theorem 4.2). Here, and throughout, for any $p \times s$ matrix a of rank $s < p$ we denote by $a_\perp$ a $p \times (p-s)$ matrix such that $a_\perp' a = 0$, and for $s \leq p$ we define $\bar{a} = a(a'a)^{-1}$.

Assumption 1. *The roots of* $\det \Pi(z) = 0$ *are either greater than one in absolute value or equal to 1, thus ruling out seasonal roots. 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 $(\beta, \Gamma'\alpha_\perp)$ has full rank, see (51), and we use this property throughout. In the statistical analysis of the model $\mathcal{H}_r^{add}$ in (1), we assume freely varying parameters. Thus, for example, α and β will be freely varying $p \times r$ matrices. For the probability analysis of the data generating process, however, we assume the conditions of Assumption 1. Thus, for instance, the true values of α and β will be of full rank, and the matrix C is well defined because $\alpha_\perp' \Gamma \beta_\perp$ has full rank when evaluated at the true values. We can then find the solution of the model equations (1) for Y_t. The solution is given by the following version of Granger's Representation Theorem,

$$Y_t = C \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{t-1} C_i^* \varepsilon_{t-i} + A_t, \tag{4}$$

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

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

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

The additive formulation has been analyzed by several authors, but each for very specific choices of deterministic terms. For example, for Z_t in model (1), Lütkepohl and Saikkonen (2000) and Saikkonen and Lütkepohl (2000a) consider a linear trend, Saikkonen and Lütkepohl (2000b) and Trenkler, Saikkonen, and Lütkepohl (2007) consider a linear trend together with an impulse dummy and a shift dummy, while Nielsen (2004, 2007) considers a linear trend together with impulse dummies. In contrast, we give a unified analysis of general deterministic terms.

1.2 The innovation formulation

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

$$\Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \tilde{\gamma} Z_t + \varepsilon_t, \tag{6}$$

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 innovation formulation. They point out that the asymptotic distribution of the test for rank contains nuisance parameters, and that these 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 innovation 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 (6) is given by, see (4),

$$X_t = C \sum_{i=1}^{t} (\varepsilon_i + \tilde{\gamma} Z_i) + \sum_{i=0}^{t-1} C_i^* (\varepsilon_{t-i} + \tilde{\gamma} Z_{t-i}) + A_t. \tag{7}$$

A model like (6) is easy to estimate using reduced rank regression, but it follows from (7) that the deterministic terms in the process 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 (6) becomes a combination of a step dummy from the first term in the process (7) and an exponentially decreasing function from the second term in (7), giving a gradual shift from one level to another. A constant in the equation (6) becomes a linear function in the process (7), 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 innovation 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). An application using various dummies, including a "volcanic function" dummy variable for modeling volcanic eruptions, is given in Model V of Pretis (2015), see 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 of $\mathcal{H}_r^{add}$ in (1), we find by applying $\Pi(L)$ on both sides of (1) that $\mathcal{H}_r^{add}$ has the alternative formulation

$$\mathcal{H}_r^{add} : \Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \gamma \Delta Z_t - \alpha\beta'\gamma Z_{t-1} - \sum_{i=1}^{k-1} \Gamma_i \gamma \Delta Z_{t-i} + \varepsilon_t. \quad (8)$$

From (8) 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 (8) for X_t, the coefficients $(\gamma, -\alpha\beta'\gamma, -\Gamma_1\gamma, \ldots, -\Gamma_{k-1}\gamma)$ all 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 innovation formulation of the deterministic term $(\Delta Z_t, Z_{t-1}, \Delta Z_{t-1}, \ldots, \Delta Z_{t-k+1})$.

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. Now consider extending 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 of ρ, γ from the extended model. Alternatively, we can use these (consistent) estimators as starting values for an iteration to find the MLE.

Extending model (8) 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 Formulation of a class of regressors

If a univariate deterministic regressor U_t has the property that it is linearly dependent on some of its differences, i.e. $\sum_{i=0}^{n} c_i \Delta^i U_t = 0$ 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 $a = \pm \mathrm{i}, \mathrm{i} = \sqrt{-1}$, 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 of their differences, but for regressors that are linearly dependent on their differences we only consider a polynomial and seasonal dummies.

For asymptotic analysis, such as proving consistency of a regression coefficient, an important property of a regressor is whether its information is divergent (in which case consistency usually follows) or bounded (in which case consistency cannot be shown). For a general regressor, U_t, the simple inequality $(\Delta^{i+1} U_t)^2 \leq 2(\Delta^i U_t)^2 + 2(\Delta^i U_{t-1})^2$ shows that if the information of $\Delta^i U_t$ is bounded, then it is also bounded for $\Delta^{i+1} U_t$. On the other hand, if the information of $\Delta^{i+1} U_t$ is divergent then so is the information of $\Delta^i U_t$. Thus, for a given deterministic regressor these considerations motivate the following definition of the order of a regressor.

Definition 1. *For a univariate deterministic regressor U_t 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 largest integer i for which the information of $\Delta^i U_t$ diverges, i.e.*

$$m = \sup\{i \geq 0 : \sum_{t=1}^{T} (\Delta^i U_t)^2 \to \infty \text{ as } T \to \infty\}. \tag{9}$$

In particular, if the information of $\Delta^i U_t$ diverges for all i we define the order to be ∞. Finally, if the information of U_t is bounded, we define the order to be $m = -1$.

Example 1. For a polynomial in t of degree m, say $P_m(t)$, we note that $\Delta^m P_m(t)$ is a constant which has diverging information, but $\Delta^{m+1} P_m(t) = 0$. Thus, the order (9) of the polynomial is equal to the degree, m. More generally, for the power function $U_t = t^a$, with $a \in \mathbb{R}$ and $a > -1/2$, the order of U_t is $m = [a + 1/2]$, where $[x]$ denotes the integer part of x. ♦

Example 2. For the impulse dummy $U_t = 1_{\{t=t_0\}}$, where $1_{\{A\}}$ denotes the indicator function for the event A, we find $\sum_{t=1}^T (\Delta^i U_t)^2 \to c_i$ for all i, so the order of U_t is $m = -1$. In this case, all differences $\Delta^i U_t$ are linearly independent. 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 $m = 1$. ♦

Example 3. For the semi-annual dummy $U_t = (-1)^t$ (orthogonalized on the constant) we have $\Delta^i U_t = (-2)^{i+1} U_t$, so that $\sum_{t=1}^T (\Delta^i U_t)^2 = 4^{i+1} T \to \infty$ for all i, and hence the order is $m = \infty$. Moreover, for the semi-annual dummy we find the linear dependence $\Delta U_t = 2U_t = M_2 U_t$. Similarly, for the quarterly dummy $U_{1t} = \mathsf{i}^t + (-1)^t + \mathsf{i}^{-t}$ (also orthogonalized on the constant) and $U_t = (U_{1t}, U_{1,t-1}, U_{1,t-2})' \in \mathbb{R}^3$, we find that $\Delta U_t = M_4 U_t$, where

$$M_4 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}. \tag{10}$$

The matrix M_4 has eigenvalues $(2, 1 + \mathsf{i}, 1 - \mathsf{i})$. In general, a seasonal dummy with s seasons corrected for the constant is given by $U_{1t} = \sum_{u=1}^{s-1} (e^{2\pi \mathsf{i} u/s})^t$, so setting $U_t = (U_{1,t}, \dots, U_{1,t-s+1})'$ we find the linear dependence $\Delta U_t = M_s U_t$, where the matrix M_s has eigenvalues $1 - e^{2\pi \mathsf{i} j/s}$, $j = 1, \dots, s-1$. ♦

The regressors considered are conveniently expressed in differences (rather than lags) since these have natural interpretations in many cases. Furthermore, as both Definition 1 and 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 $U_{1t}, \dots, U_{qt}$ be linearly independent regressors with orders $m_v < \infty$ for $v = 1, \dots, q$. Assume further that $\{\Delta^i U_{vt}, i \geq 0\}$ are either linearly independent for all $i \geq 0$, or (for a polynomial) equal to zero for $i > m_v$. Let $U_{se,t}$ be an $(s-1)$-dimensional seasonal dummy variable orthogonalized to the constant term. We define the multivariate q-dimensional regressor $U_t = (U_{1t}, \dots, U_{qt})'$ and consider the regressor defined as*

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

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

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

such that

$$\gamma Z_t = \sum_{i=0}^n \gamma^i \Delta^i U_t + \gamma^{se} U_{se,t} = \sum_{v=1}^q \sum_{i=0}^n \gamma_v^i \Delta^i U_{vt} + \gamma^{se} U_{se,t}. \tag{11}$$

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 dynamic parameters in π. In particular, see (1),

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

The deterministic term in the additive model equation, see (8) and (11), is then

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

where we have introduced the coefficients $\Upsilon_0, \Upsilon_1, \ldots, \Upsilon_{n+k}, \Upsilon_{se}$ depending on (π,γ) and given by

$$\Upsilon_i = \Upsilon_i(\pi,\gamma) = \sum_{j=\max\{0,i-n\}}^{\min\{i,k\}}\Phi_j(\pi)\gamma^{i-j}, 0 \le i \le n+k, \Upsilon_{se} = \sum_{i=0}^{k}\Phi_i(\pi)\gamma^{se}M_s^i. \tag{14}$$

In particular, we find

$$\Upsilon_0(\pi,\gamma) = \Phi_0(\pi)\gamma^0 = -\alpha\beta'\gamma^0 \text{ and } \Upsilon_1(\pi,\gamma) = \Phi_0(\pi)\gamma^1 + \Phi_1(\pi)\gamma^0 = -\alpha\beta'\gamma^1 + (\alpha\beta' + \Gamma)\gamma^0.$$

We note that Υ_0 is proportional to α, and define the parameter $\rho' = \bar{\alpha}'\Upsilon_0 = \bar{\alpha}'\Phi_0\gamma^0 = -\beta'\gamma^0$. It is then clear from (14) that, for given values of the dynamic parameter π, the parameters ρ and $\Upsilon = (\Upsilon_1, \ldots, \Upsilon_{n+k}, \Upsilon_{se})$ are linear functions of γ, and for given β also linear functions of π. In Theorem 1 we next give an algorithm for recovering the parameter γ as a linear function of the parameters $\Upsilon_0, \Upsilon_1, \ldots, \Upsilon_n, \alpha_\perp'\Gamma\Upsilon_{n+1}, \Upsilon_{se}$, for given values of π.

Theorem 1. *Let Assumption 1 be satisfied and consider the deterministic term in the additive model, $\Pi(L)\gamma Z_t$, and the parameter functions $\Upsilon_0, \Upsilon_1, \ldots, \Upsilon_{n+k}, \Upsilon_{se}$, see (13) and (14). Then, for $i = 0, \ldots, n$,*

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

which can be solved for γ^i because $(\beta, \Gamma'\alpha_\perp)$ has full rank. Moreover,

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

The relations (15) show that, for a given value of π, the parameters $(\gamma^0, \ldots, \gamma^{i-1}, \beta'\gamma^i)$ can be found recursively as an invertible linear function of $(\rho, \Upsilon_1, \ldots, \Upsilon_i)$ because $(\beta, \Gamma'\alpha_\perp)$ has full rank, see Assumption 1. In particular, we can find γ^0 from

$$\rho' = -\beta'\gamma^0 = \bar{\alpha}'\Upsilon_0 \text{ and } \alpha_\perp'\Gamma\gamma^0 = \alpha_\perp'\Upsilon_1. \tag{17}$$

Note that the coefficients $\beta'\Upsilon_{n+1}, \Upsilon_{n+2}\ldots, \Upsilon_{n+k}$ are not needed to recover the coefficients in γ.

2.3 The extended model

Corresponding to the additive model $\mathcal{H}_r^{add}$ with regressors γZ_t, we define the extended model $\mathcal{H}_r^{ext}$ using (8) and the coefficients in (14) by

$$\mathcal{H}_r^{ext} : \Delta X_t = \alpha(\beta' X_{t-1} + \rho' U_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=1}^{n+k} \Upsilon_i \Delta^i U_t + \Upsilon_{se} U_{se,t} + \varepsilon_t, \tag{18}$$

where the parameters in $\xi = (\pi, \rho, \Upsilon)$ are freely varying. Thus ρ, Υ^1 are not in general functions of π, γ, but the additive model $\mathcal{H}_r^{add}$ in (8) is a submodel of the extended model $\mathcal{H}_r^{ext}$ in (18). For the true values, the relations (14) express the true value of the extended model parameter, ξ_0, as a function of the true value of the additive model parameter, $\lambda_0 = (\pi_0, \gamma_0)$.

2.4 Comparison of the additive and extended models

We compare the additive and extended models in a simple case with lag length $k = 1$ and a univariate regressor, U_t, with order $m = 1$, which can illustrate the role of the order and the choice of n. The additive and extended models have the same dynamic parameters, (α, β), and only differ in the choice of regressors and their coefficients. Because $m = 1$, U_t and ΔU_t have diverging information, while $\Delta^2 U_t$ has bounded information, see Definition 1. First suppose we choose the deterministic term $Z_t = U_t$ in the additive model, i.e. $n = 0$. The additive model and associated extended model are given by

$$\mathcal{H}_r^{add} : \Delta X_t = \alpha(\beta' X_{t-1} - \beta' \gamma U_t) + (\alpha\beta' + I_p)\gamma \Delta U_t + \varepsilon_t, \tag{19}$$

$$\mathcal{H}_r^{ext} : \Delta X_t = \alpha(\beta' X_{t-1} + \rho' U_t) + \Upsilon_1 \Delta U_t + \varepsilon_t. \tag{20}$$

In this case the deterministic term in $\mathcal{H}_r^{add}$ has p parameters, γ, whereas $\mathcal{H}_r^{ext}$ has $r + p$ parameters, (ρ, Υ_1). Thus, the additive model is a genuine submodel of the extended model. An analysis of inference in this model is given in Theorem 9.

If instead we choose the deterministic term $\gamma Z = \gamma^0 U_t + \gamma^1 \Delta U_t$, i.e. $n = 1$, then we find

$$\mathcal{H}_r^{add} : \Delta X_t = \alpha(\beta' X_{t-1} - \beta' \gamma^0 U_t) + ((\alpha\beta' + I_p)\gamma^0 - \alpha\beta'\gamma^1)\Delta U_t + (\alpha\beta' + I_p)\gamma^1 \Delta^2 U_t + \varepsilon_t, \tag{21}$$

$$\mathcal{H}_r^{ext} : \Delta X_t = \alpha(\beta' X_{t-1} + \rho' U_t) + \Upsilon_1 \Delta U_t + \Upsilon_2 \Delta^2 U_t + \varepsilon_t. \tag{22}$$

Now the extended model has $r + 2p$ parameters, $(\rho, \Upsilon_1, \Upsilon_2)$, but Υ_2 cannot be estimated consistently, because $\Delta^2 U_t$ has bounded information. This leaves $p + r$ parameters, (ρ, Υ_1), for $\mathcal{H}_r^{ext}$. The additive model has parameters γ^0 and γ^1, but because $\Delta^2 U_t$ has bounded information, the only large-sample information comes from the coefficients $\beta'\gamma^0$ and $(\alpha\beta' + I_p)\gamma^0 - \alpha\beta'\gamma^1$, in the sense that only these coefficients can be estimated consistently. We can obviously recover $\beta'\gamma^0$ from the first coefficient and, using Theorem 1, we can find $\alpha'_\perp\gamma^0$ as well as $\beta'\gamma^1$ from the second coefficient. Thus, for large-sample inference, the additive model also has $p + r$ parameters, $(\gamma^0, \beta'\gamma^1)$, which are in one-to-one correspondence with the parameters (ρ, Υ_1) from the extended model. This is an example of an additive model, where n has been chosen greater than or equal to $m = 1$. In this case the models $\mathcal{H}_r^{add}$ and $\mathcal{H}_r^{ext}$ are reparametrizations of each other, if we remove regressors with bounded information, and in that sense the models are asymptotically equal, see Theorem 4. An analysis of inference in this model is given in Theorem 8.

In either case, note that one can include the regressors with bounded information in the estimation, but they automatically disappear in the asymptotic analysis, as we shall show in Lemma 1 and Theorem 4.

In general $\mathcal{H}_r^{add}$ is a submodel of $\mathcal{H}_r^{ext}$, but there is a special case where the two models are the same, as given in the next theorem.

Theorem 2. *Consider the polynomial regressors* $f_i(t) = (t+i)\dots(t+1)/i!$ *for* $1 \leq i \leq m, t \geq 0$ *and* $f_0(t) = 1$ *for* $t \geq 0$. *Then* $\Delta f_i(t) = f_{i-1}(t)$ *so that the additive model has deterministic term* $\gamma Z_t = \sum_{i=0}^{m} \gamma^i f_{m-i}(t)$ *and the corresponding extended model has deterministic term* $\alpha\rho' f_m(t) + \sum_{i=1}^{m} \Upsilon_i f_{m-i}(t)$, *see (1) and (18). If Assumption 1 is satisfied, then the additive model is a reparametrization of the extended model.*

The regressor $Z_t = (f_m(t), f_{m-1}(t), \dots, f_0(t))'$ in Theorem 2 is equivalent to the more common regressor $Z_t = (t^m, \dots, 1)'$, in the sense that they span the same space of polynomials. The form of Z_t shows that we have chosen $n = m$ in this case. The deterministic term thus contains $pm + r$ identified parameters, $\gamma^0, \dots, \gamma^{m-1}, \beta'\gamma^m$, in the additive model, and the same number of identified parameters in the form of $\rho, \Upsilon_1, \dots, \Upsilon_m$ in the extended model. Theorem 1 shows that $\gamma^0, \dots, \gamma^{m-1}, \beta'\gamma^m$ from the additive model can be determined uniquely from $\rho, \Upsilon_1, \dots, \Upsilon_m$ from the extended model for a given value of π. Thus, for this choice of regressors, the additive model parametrized by $\lambda = (\pi, \gamma^0, \dots, \gamma^{m-1}, \beta'\gamma^m)$ is the same as the extended model parametrized by $\xi = (\pi, \rho, \Upsilon_1, \dots, \Upsilon_m)$. With the choice of Z_t in Theorem 2, the models with $m = 0$ and $m = 1$ are analyzed in Johansen (1996, Chapter 5) as $H^*(r), H_1^*(r)$. Note that the result in Theorem 2 still holds if Z_t were enlarged to include also a seasonal dummy.

In general, of course, the additive model is not a reparametrization of the extended model. In order to derive the simple result that the additive model and the extended model with regressors with bounded information removed, are reparametrizations, we will need to make the next assumption.

Assumption 2. *For the regressor* $Z_t = (U_t', \dots, \Delta^n U_t', U_{se,t}')'$ *we choose* $n \geq \max_{1\leq v\leq q} m_v$, *where* m_v *is the order of* U_{vt}, $v = 1, \dots, q$, *see Definition 2.*

The important condition in Assumption 2 is that, for each regressor in the additive model, one should also include its differences, as long as these differences have diverging information. As an illustration of a situation with $n = 0$ but $m = 1$, in which Assumption 2 is violated, consider the following example.

Example 4. Consider the model $X_t = Y_t + \gamma(t - t_0)^+$ and $\Delta Y_t = \alpha\beta' Y_{t-1} + \varepsilon_t$, where not including the step dummy associated with the broken trend function has the effect of enforcing continuity (in t) of the deterministic term. From (19) and (20) we find the additive and extended models,

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

We note that in this example, Assumption 2 would be satisfied by including a step dummy, $1_{\{t\geq t_0+1\}} = \Delta(t - t_0)^+$, in the additive model formulation. This is illustrated as follows.

Example 5. Continuation of Example 4. Suppose we include the missing step dummy, $1_{\{t\geq t_0+1\}} = \Delta(t-t_0)^+$, in the additive model formulation such that $X_t = Y_t + \gamma^0(t-t_0)^+ + \gamma^1 1_{\{t\geq t_0+1\}}$ giving

$$\begin{aligned}\mathcal{H}_r^{add}: \Delta X_t =& \alpha(\beta' X_{t-1} - \beta'\gamma^0(t-t_0)^+) + ((\alpha\beta' + I_p)\gamma^0 - \alpha\beta'\gamma^1)1_{\{t\geq t_0+1\}} \\ &+ (\alpha\beta' + I_p)\gamma^1 1_{\{t=t_0+1\}} + \varepsilon_t, \\ \mathcal{H}_r^{ext}: \Delta X_t =& \alpha(\beta' X_{t-1} + \rho(t-t_0)^+) + \upsilon_1 1_{\{t\geq t_0+1\}} + \upsilon_2 1_{\{t=t_0+1\}} + \varepsilon_t,\end{aligned}$$

see (21) and (22). With this slightly larger additive model we have $n = 1$ and $m = 1$. 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 2 is now satisfied. ♦

For the general case, we next discuss identification and estimation of the parameters in the additive model, in the situation where we have included a polynomial regressor to illustrate what happens with identification and estimation of the constant term.

2.5 Identification of the parameters in the extended and additive models

Consider the parameter $\xi = (\pi, \rho, \Upsilon)$ in the extended model (18), where a polynomial regressor U_{1t} is included and U_{1t} is of order m_1. We assume that the zero regressors of the polynomial, i.e. $\Delta^i U_{1t} = 0$ for $i > m_1$, have been removed together with their coefficients, so that the remaining regressors are linearly independent (Definition 2). Therefore ξ is identified, because if the likelihood functions for parameters ξ_1 and ξ_2 are the same, then in particular for the deterministic terms,

$$\alpha_1\rho_1' U_t + \sum_{i=1}^{n+k}\Upsilon_{1i}\Delta^i U_t + \Upsilon_{1,se}U_{se,t} = \alpha_2\rho_2' U_t + \sum_{i=1}^{n+k}\Upsilon_{2i}\Delta^i U_t + \Upsilon_{2,se}U_{se,t} \text{ for all } t,$$

such that $\xi_1 = \xi_2$, because of linear independence of the retained regressors, 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 (8) as a submodel of the extended model (18). Identification of the additive model is a consequence of the following result, which is based on Theorem 1. The result is formulated for the additive model with a polynomial regressor to illustrate what happens to identification of the constant term, which generates a zero regressor when differenced.

Theorem 3. *Let Assumption 1 be satisfied. Let $\lambda = (\pi, \gamma)$ be the parameters in the additive model (8), 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 $\xi = \xi(\lambda) = (\pi, \rho, \Upsilon)$, where ρ, Υ are defined by (14) and $\rho' = \bar{\alpha}'\Upsilon_0$, and assume the coefficients $\Upsilon_{i1}, i > m_1$, corresponding to the polynomial have been removed. Then, for any set of parameters λ_0 and λ_h, $h \to 0$, we find*

$$\xi(\lambda_h) \to \xi(\lambda_0) \text{ as } h \to 0 \text{ implies } \lambda_h \to \lambda_0, \tag{23}$$

except for the constant term with coefficient $\gamma_{1,h}^{m_1}$ if $n \geq m_1$, where 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 3 because if $\xi(\lambda_1) = \xi(\lambda_0)$ then, choosing $\lambda_h = \lambda_1$, we find from (23) that $\lambda_1 = \lambda_0$. Thus, a special case of Theorem 3 implies identification of the parameters of the additive model in the usual sense.

However, in anticipation of our proof of consistency, Theorem 3 proves the more general result that ξ depends continuously on the parameter λ, which one could call "continuous identification". The function $\xi(\lambda)$ is clearly a continuous function of all parameters. If Λ denotes the parameter space for λ and we let $\xi(\Lambda)$ denote the image of Λ, then Theorem 1 shows that the function ξ restricted to $\xi(\Lambda)$ is invertible. What is shown in Theorem 3 is that this inverse function is continuous on $\xi(\Lambda)$.

The result in Theorem 3 thus 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.

2.6 Estimation of the parameters in the extended and additive models

For estimation 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 (18) 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).

The additive model (8) has no simple closed-form estimation algorithm, but one can use a numerical optimization algorithm to maximize the likelihood function, using that the model is 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 Theorem 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 distribution, both for the additive model and the extended model.

3.1 Regressors with bounded information

In the application of the models it is useful to allow, for instance, impulse dummies like $U_{vt} = 1_{(t=t_0)}$ as a regressor to account for an outlier. However, for the asymptotic analysis, regressors with bounded information will not give consistent estimation or asymptotically Gaussian inference for 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 for $i > m_v$, see Definition 1. We next prove a result that allows us to disregard regressors with bounded information in the asymptotic analysis of the model, in the sense that these regressors have no influence on asymptotic inference for the remaining parameters.

Lemma 1. *Let Z_{1t} be stochastic or deterministic with diverging information, $\sum_{i=1}^T Z_{1t}^2 \overset{\mathrm{P}}{\to} \infty$, and let Z_{2t} be deterministic with bounded positive information, $0 < \sum_{t=1}^T Z_{2t}^2 \leq c$. Then*

$$\frac{\sum_{t=1}^T Z_{1t} Z_{2t}}{(\sum_{t=1}^T Z_{2t}^2)^{1/2} (\sum_{t=1}^T Z_{1t}^2)^{1/2}} \overset{\mathrm{P}}{\to} 0.$$

A consequence of Lemma 1 is that the limit of the information matrix normalized by its diagonal elements will be asymptotically block diagonal, with one block corresponding to regressors with bounded information. This is used to prove that the latter type of regressors do not contribute to the asymptotic distribution of the remaining parameters. To this end, we now define the additive and extended models where regressors with bounded information have been removed, and subsequently we give the result.

Definition 3. *Define $\mathcal{H}_r^{add*}$ as the additive "core" model, given by $\mathcal{H}_r^{add}$ in (8), but where regressors with bounded information have been removed. Similarly define $\mathcal{H}_r^{ext*}$ as the extended "core" model, given by $\mathcal{H}_r^{ext}$ in (18), but where regressors with bounded information have been removed.*

Theorem 4. *The asymptotic distribution of the MLEs in $\mathcal{H}_r^{add*}$ is the same as the asymptotic distribution of the MLEs of the same parameters in $\mathcal{H}_r^{add}$, and the asymptotic distribution of the MLEs in $\mathcal{H}_r^{ext*}$ is the same as the asymptotic distribution of the MLEs of the same parameters in $\mathcal{H}_r^{ext}$. Moreover, if Assumptions 1 and 2 are satisfied, then the models $\mathcal{H}_r^{ext*}$ and $\mathcal{H}_r^{add*}$ are reparametrizations of each other.*

There are two important results in Theorem 4. First, the asymptotic distributions of the MLEs in the core models are the same as the asymptotic distributions of the MLEs of the same parameters in the models that include the regressors with bounded information. Consequently, we will therefore assume in the asymptotic analysis that all such regressors and their coefficients have been removed. Second, under Assumptions 1 and 2, Theorem 4 shows that the two core models, $\mathcal{H}_r^{ext*}$ and $\mathcal{H}_r^{add*}$, are reparametrizations of each other. That is, in general $\mathcal{H}_r^{ext}$ and $\mathcal{H}_r^{add}$ are not reparametrizations as for polynomials, see Theorem 2, but Assumption 2 is the important condition that allows us to establish that, asymptotically, a result that parallels Theorem 2 holds for the core models.

Example 6. Continuation of Examples 4 and 5. It is seen that the two models in Example 5 are not reparametrizations as for polynomials, see Theorem 2, but the coefficient υ_2 is associated with a regressor with information $\sum_{t=1}^{T} 1_{\{t=t_0+1\}}^2 = 1$, and hence does not contribute to the asymptotic analysis (Lemma 1). 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 2 is now satisfied and the two models in Example 5 are asymptotically equivalent in the sense that their respective core models are reparametrizations of each other, see Theorem 4. ♦

3.2 Partition and normalization of regressors with diverging information

In the following we consider only the core models, $\mathcal{H}_r^{ext*}$ and $\mathcal{H}_r^{add*}$, based on the arguments in Theorem 4. That is, we remove regressors U_{vt} with bounded information and assume, without loss of generality, that all components U_{vt} have $m_v \geq 0$, and we discard the regressors $\Delta^i U_{vt}$ with $i > m_v$. We then find that the deterministic term in the extended model (18),

$$\alpha \sum_{v=1}^{q} \rho_v' U_{vt} + \sum_{i=1}^{n+k} \sum_{v=1}^{q} \Upsilon_{iv} \Delta^i U_{vt} + \Upsilon_{se} U_{se,t}, \tag{24}$$

can be replaced—without changing the results of the asymptotic analysis—by

$$\alpha \rho' Z_{0t} + \Upsilon^1 Z_{1t}, \tag{25}$$

where $Z_{0t} = U_t$ and Z_{1t} are the regressors with diverging information including the seasonal dummies,

$$Z_{1t} = (\Delta^i U_{vt}, 1 \leq i \leq \min\{n+k, m_v\}, 1 \leq v \leq q; U'_{se,t})', \tag{26}$$

and where the corresponding freely varying coefficients are

$$\rho' = -\beta'\gamma^0 \text{ and } \Upsilon^1 = (\Upsilon_{iv}, 1 \leq i \leq \min\{n+k, m_v\}, 1 \leq v \leq q; \Upsilon_{se}). \tag{27}$$

We note that Z_{1t} may be empty, in which case the remainder of the analysis is easily simplified accordingly. We also note that the true values ρ_0 and Υ_0^1 are both functions of (π_0, γ_0), see (14).

The asymptotic analysis is based on the behaviour of suitable product moments. We therefore 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 the residuals of U_t corrected for W_t,

$$(U_t | W_t) = U_t - \langle U, W \rangle_T \langle W, W \rangle_T^{-1} W_t.$$

Product moments of residuals are denoted

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

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

Next, for the asymptotic analysis we need the following normalizations of the regressors and a mild condition to rule out asymptotically multicollinear regressors. For a regressor with diverging information, i.e. $\Delta^i U_{vt}$ with $i \leq m_v$, we introduce the normalization M_{Tiv}. The normalizations of Z_{0t} and Z_{1t} are then given by the diagonal matrices

$$\begin{aligned} N_{T0} &= \text{diag}(M_{T0v}, 1 \leq v \leq q), \\ N_{T1} &= \text{diag}(M_{Tiv}, 1 \leq i \leq m_v, 1 \leq v \leq q; \iota'_{s-1}), \end{aligned}$$

where ι_{s-1} is an $(s-1)$-vector of ones, because the $s-1$ seasonal dummies need no normalization. This defines the normalized regressors

$$Z_{0Tt} = N_{T0}^{-1} Z_{0t} \text{ and } Z_{1Tt} = N_{T1}^{-1} Z_{1t}. \tag{28}$$

Assumption 3. *The normalizations satisfy $M_{Tiv}^{-1} T^{-1/2} \to 0$ and $M_{Tiv}^{-1} M_{T,i+1,v} \to 0$ and the asymptotic information matrix for the normalized regressors is nonsingular, i.e. satisfies*

$$\left\langle \begin{pmatrix} Z_{0T} \\ Z_{1T} \end{pmatrix}, \begin{pmatrix} Z_{0T} \\ Z_{1T} \end{pmatrix} \right\rangle_T \to \left\langle \begin{pmatrix} Z_0 \\ Z_1 \end{pmatrix}, \begin{pmatrix} Z_0 \\ Z_1 \end{pmatrix} \right\rangle > 0. \tag{29}$$

Example 7. 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 $U_t = (t + 1_{\{t \geq t_0+1\}}, t + 1_{\{t \geq t_0-1\}})'$ and suppose $n = m = 1$ such that $Z_{0t} = U_t, Z_{1t} = \Delta U_t$, see Definition 2 and (26). We normalize by $M_{T0v} = T$ and $M_{T1v} = 1$

and find $Z_{0T[T\nu]} \to (\nu,\nu)'$ and $Z_{1T[T\nu]} \to (1,1)'$, such that the limit in (29) is singular. In this case, one could choose instead $U_t = (t+1_{\{t\geq t_0+1\}}, 1_{\{t\geq t_0-1\}} - 1_{\{t\geq t_0+1\}})'$, which spans the same space, but $T^{-1}U_{[T\nu]} \to (\nu,0)'$ and $\Delta U_{[T\nu]} \to (1,0)'$, such that we can discard the second component and find $Z_{0T[T\nu]} \to \nu$, $Z_{1T[T\nu]} \to 1$, which gives rise to consistent estimation with a non-singular asymptotic information matrix. ♦

Finally, for a sequence ε_t, which is i.i.d. $(0,\Omega)$, we define the Brownian motion W_ε as the weak limit of the partial sum of ε_t; that is, for $S_t = \sum_{i=1}^t \varepsilon_i$ we define W_ε from

$$T^{-1/2}S_{[T\nu]} = T^{-1/2}\sum_{t=1}^{[T\nu]}\varepsilon_i \xrightarrow{\text{D}} W_\varepsilon(\nu). \tag{30}$$

Of course, W_ε can be considered as $W_\varepsilon = \Omega^{1/2}B$, where B is standard Brownian motion, but we find that using W_ε is a simpler notation. For the asymptotic analysis we make the following high-level assumption, for which primitive sufficient conditions are well-known.

Assumption 4. *The following limits as $T\to\infty$ exist and the convergences hold jointly,*

$$\begin{aligned}
T^{1/2}\langle Z_{jT},\varepsilon\rangle_T &= T^{-1/2}N_{Tj}^{-1}\textstyle\sum_{t=1}^T Z_{jt}\varepsilon_t' &&\xrightarrow{\text{D}} \langle Z_j,\varepsilon\rangle \ \text{for } j=0,1,\\
T^{-1/2}\langle Z_{jT},S_{t-1}\rangle_T &= T^{-3/2}N_{Tj}^{-1}\textstyle\sum_{t=1}^T Z_{jt}S_{t-1}' &&\xrightarrow{\text{D}} \langle Z_j,W_\varepsilon\rangle \ \text{for } j=0,1,\\
T^{-1/2}\langle S_{t-1},\varepsilon\rangle_T &= T^{-3/2}\textstyle\sum_{t=1}^T S_{t-1}\varepsilon_t' &&\xrightarrow{\text{D}} \textstyle\int_0^1 W_\varepsilon(dW_\varepsilon)' = \langle W_\varepsilon,\varepsilon\rangle.
\end{aligned}$$

Again, we use $\langle Z_j,\varepsilon\rangle$, 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 limits in Assumption 4 are given next.

Example 8. Let $U_t = (t,(t-t_0)^+)'$ with $\Delta U_t = (1,1_{\{t\geq t_0+1\}})'$. Then $M_{T0v}=T$ and $M_{T1v}=1$ and we note that $M_{Tiv}^{-1}T^{-1/2}\to 0$ and $M_{T0v}^{-1}M_{T1v}\to 0$, reflecting that the order of the regressor in this case decreases when differenced. We define

$$\begin{aligned}
u(\nu) &= \lim_{T\to\infty} U_{T,[T\nu]} = \lim_{T\to\infty} N_{T0}^{-1}U_{[T\nu]} = (\nu,(\nu-\nu_0)^+)',\\
\dot u(\nu) &= \lim_{T\to\infty} \Delta U_{T,[T\nu]} = \lim_{T\to\infty} N_{T1}^{-1}\Delta U_{[T\nu]} = (1,1_{\{\nu>\nu_0\}})'.
\end{aligned}$$

For this example we find the limits, as $T\to\infty$,

$$\begin{aligned}
\langle U_T,\Delta U_T\rangle_T &= T^{-1}\textstyle\sum_{t=1}^T(T^{-1}U_t)(\Delta U_t)' &&\to \textstyle\int_0^1 u(\nu)\dot u(\nu)'d\nu = \langle U,\Delta U\rangle,\\
T^{1/2}\langle U_T,\varepsilon\rangle_T &= T^{-1/2}\textstyle\sum_{t=1}^T T^{-1}U_t\varepsilon_t' &&\xrightarrow{\text{D}} \textstyle\int_0^1 u(\nu)dW_\varepsilon(\nu)' = \langle U,\varepsilon\rangle,\\
T^{-1/2}\langle U_T,S_{t-1}\rangle_T &= T^{-3/2}\textstyle\sum_{t=1}^T T^{-1}U_tS_{t-1}' &&\xrightarrow{\text{D}} \textstyle\int_0^1 u(\nu)W_\varepsilon(\nu)'d\nu = \langle U,W_\varepsilon\rangle.
\end{aligned}$$
♦

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

Example 9. Let $U_{se,t} = (-1)^t$ be a seasonal dummy variable. Then, as $T \to \infty$,

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

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

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. ♦

3.3 Consistency of parameter estimators

Because we have eliminated regressors with bounded information by the above arguments, see in particular Theorem 4, from this point onwards the asymptotic analysis will focus on the additive and extended core models, $\mathcal{H}_r^{add*}$ and $\mathcal{H}_r^{ext*}$, see (8), 18), and Definition 3. The parameters of the additive core model, $\mathcal{H}_r^{add*}$, are given by $\lambda = (\pi, \gamma)$ with true value $\lambda_0 = (\pi_0, \gamma_0)$, and the parameters of the extended core model, $\mathcal{H}_r^{ext*}$, are given by $\xi = (\pi, \rho, \Upsilon^1)$ with true value $\xi_0 = (\pi_0, \rho_0, \Upsilon_0^1) = (\pi_0, -\beta_0'\gamma_0, \Upsilon^1(\pi_0, \gamma_0))$.

Let θ denote a parametrization of the conditional mean, e.g. λ or ξ. Then, for this parametrization, the negative (quasi-) log-likelihood function is, apart from a constant term,

$$L(\theta) = -\log L_T(\theta, \Omega = \Omega_0) = \frac{1}{2} \operatorname{tr}\{\Omega_0^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta) \varepsilon_t(\theta)'\}, \tag{31}$$

where we have set $\Omega = \Omega_0$ (without loss of generality for asymptotic inference on the remaining parameters) and $\varepsilon_t(\theta)$ are the residuals, which are defined for the additive and extended (core) models as

$$\mathcal{H}_r^{add*} : \varepsilon_t(\lambda) = \Delta X_t - \alpha(\beta' X_{t-1} - \beta' \gamma^0 Z_{0t}) - \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} - \Upsilon^1(\pi, \gamma) Z_t, \tag{32}$$

$$\mathcal{H}_r^{ext*} : \varepsilon_t(\xi) = \Delta X_t - \alpha(\beta' X_{t-1} + \rho' Z_{0t}) - \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} - \Upsilon^1 Z_{1t}, \tag{33}$$

see (1) and (18). The MLEs of λ and ξ are then defined as $\breve{\lambda} = \arg\min_\lambda L(\lambda)$ and $\hat{\xi} = \arg\min_\xi L(\xi)$, respectively. The latter can be obtained by reduced rank regression, but the former requires numerical optimization, see Section 2.6.

To prove consistency, we use the fact that both the additive and extended models can be expressed as nonlinear submodels of a linear regression model. The Gaussian log-likelihood function in the linear regression model is quadratic and therefore the level curves are ellipses. This means that, for any ellipse, the log-likelihood is smaller outside the ellipse than all values inside the ellipse. This simple fact can be used to prove consistency for a nonlinear submodel, as in the following result from Johansen (2006, Lemma 16, p. 114), which we will apply.

Lemma 2. *Consider the regression model* $y_t = \varsigma' z_t + \varepsilon_t$*, where* ε_t *is i.i.d.* $(0, \Omega_0)$*, with stochastic or deterministic regressors, and let* $\varsigma = \varsigma(\tau)$ *be a continuously identified parametrization of a submodel. We define the estimator* $\hat{\tau}$ *as the minimizer of*

$$\mathrm{tr}\{\Omega_0^{-1} \sum_{t=1}^{T} (y_t - \varsigma(\tau)' z_t)(y_t - \varsigma(\tau)' z_t)'\}.$$

If the information diverges in probability for all components of z_t*, that is,*

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

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 core models thus follows from Theorems 3 and 4 because we only include regressors with divergent information.

Theorem 5. *Suppose Assumptions 1 and 3 are satisfied. In the extended core model* $\mathcal{H}_r^{ext*}$*, see Definition 3 and (18), with parameter* ξ*, the MLE* $\hat{\xi}$ *exists with probability converging to one and* $\hat{\xi} \overset{\mathrm{P}}{\to} \xi_0$*. Similarly, in the additive core model* $\mathcal{H}_r^{add*}$*, see Definition 3 and (8), with parameter* λ*, the MLE* $\breve{\lambda}$ *exists with probability converging to one and* $\breve{\lambda} \overset{\mathrm{P}}{\to} \lambda_0$*.*

3.4 Asymptotic distribution of estimators

Letting $\Pi_0(L)$ denote the characteristic polynomial with the true values π_0 inserted, we first note that the residuals (33) for the extended model are

$$\begin{aligned} \varepsilon_t(\xi) &= \Pi(L) X_t - \alpha \rho' Z_{0t} - \Upsilon^1 Z_{1t} = \Pi(L) Y_t + \Pi(L) \gamma_0 Z_t - \alpha \rho' Z_{0t} - \Upsilon^1 Z_{1t} \\ &= (\Pi(L) - \Pi_0(L)) Y_t + \Pi(L) \gamma_0 Z_t - \alpha \rho' Z_{0t} - \Upsilon^1 Z_{1t} + \varepsilon_t. \end{aligned} \tag{34}$$

The deterministic term in (34) is

$$\Pi(L) \gamma_0 Z_t - \alpha \rho' Z_{0t} - \Upsilon^1 Z_{1t} = -\alpha (\beta' \gamma_0^0 + \rho') Z_{0t} - (\Upsilon^1 - \Upsilon^1(\pi, \gamma_0)) Z_{1t}. \tag{35}$$

The stochastic term in the residuals in (34) is, using the normalization $\beta = \beta_0 + \beta_{0\perp} \bar{\beta}_{0\perp}' (\beta - \beta_0)$,

$$(\Pi(L) - \Pi_0(L)) Y_t = -(\alpha - \alpha_0) \beta_0' Y_{t-1} - \alpha (\beta - \beta_0)' \bar{\beta}_{0\perp} \beta_{0\perp}' Y_{t-1} - \sum_{i=1}^{k-1} (\Gamma_i - \Gamma_{i,0}) \Delta Y_{t-i}, \tag{36}$$

see (2), where we note in particular that, from (4) and (30),

$$T^{-1/2} \beta_{0\perp}' Y_{[T\nu]} = T^{-1/2} \beta_{0\perp}' C_0 S_{[T\nu]} + o_P(1) \overset{\mathsf{D}}{\to} \beta_{0\perp}' C_0 W_\varepsilon(\nu)$$

as $T \to \infty$.

We simplify the notation by defining new parameters to account for the parameters that actually appear in (35) and (36) and to account for different normalizations. After deriving their asymptotic distributions, we use those to derive the asymptotic distributions of the MLEs of the extended model parameters $\xi = (\pi, \rho, \Upsilon^1)$ and the additive model parameters $\lambda = (\pi, \gamma)$. Thus, a convenient device for deriving the asymptotic distributions of $\hat{\xi}$ and $\breve{\lambda}$ is to define

$$G_{Tt} = \begin{pmatrix} T^{-1/2}\beta_{0\perp}' Y_{t-1} \\ Z_{0Tt} \end{pmatrix}, \; \zeta = \begin{pmatrix} T^{1/2}\bar{\beta}_{0\perp}'(\beta - \beta_0) \\ N_{T0}(\gamma_0^{0\prime}\beta + \rho) \end{pmatrix}, \text{ and } \upsilon = (\Upsilon^1 - \Upsilon^1(\pi, \gamma_0))N_{T1}, \tag{37}$$

as well as

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

with $\Sigma_{stat} = Var(Y_t^*)$ and

$$\Upsilon^1(\pi, \gamma_0) = (\Upsilon_{iv}(\pi, \gamma_0), 1 \leq i \leq \min\{n + k, m_v\}; \Upsilon_{se}(\pi, \gamma_0)). \tag{39}$$

These definitions yield a simple expression for the residuals,

$$\varepsilon_t(\alpha^*, \zeta, \upsilon) = -\alpha^* Y_t^* - \alpha\zeta' G_{Tt} - \upsilon Z_{1Tt} + \varepsilon_t, \tag{40}$$

and it is clear that there is a simple one-to-one relation between the parameters $\xi = (\pi, \rho, \Upsilon^1)$ and $(\alpha^*, \zeta, \upsilon)$ such that $(\alpha^*, \zeta, \upsilon)$ are freely varying parameters. By an analysis of the likelihood function (31) corresponding to $\varepsilon_t(\alpha^*, \zeta, \upsilon)$ in (40), we can prove the following result, from which we will subsequently derive the asymptotic distribution of the estimators in the extended and additive models.

Theorem 6. *Suppose Assumptions 1, 3, and 4 are satisfied and that the data is generated by (1) with $\lambda = \lambda_0$. We consider the statistical model $\mathcal{H}_r^{ext*}$ and the parameters $(\alpha^*, \zeta, \upsilon)$, see (37) and (38). Then the asymptotic distributions of the MLEs $(\hat{\alpha}^*, \hat{\zeta}, \hat{\upsilon})$ based on (31) and (40) are given by*

$$T^{1/2}\hat{\alpha}^* \xrightarrow{\text{D}} N_{p\times(r+(k-1)q)}\left(0, \Sigma_{stat}^{-1} \otimes \Omega_0\right), \tag{41}$$

$$T^{1/2}\hat{\zeta} \xrightarrow{\text{D}} -\langle G, G|Z_1\rangle^{-1}\langle G, \varepsilon_\alpha|Z_1\rangle, \tag{42}$$

$$T^{1/2}\hat{\upsilon} \xrightarrow{\text{D}} -\langle \varepsilon, Z_1\rangle\langle Z_1, Z_1\rangle^{-1} + \alpha_0\langle \varepsilon_\alpha, G|Z_1\rangle\langle G, G|Z_1\rangle^{-1}\langle G, Z_1\rangle\langle Z_1, Z_1\rangle^{-1}, \tag{43}$$

where the convergences hold jointly and $\varepsilon_{\alpha,t} = (\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\varepsilon_t$. Furthermore, the distribution (41) is asymptotically independent of the distributions (42) and (43).

The distribution in (42) 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 (43) is not MG, although we can obtain asymptotic Gaussianity for the linear combinations $\hat{\alpha}_\perp'\hat{\upsilon} = \hat{\alpha}_\perp'(\hat{\Upsilon}^1 - \Upsilon^1(\hat{\pi}, \gamma_0))N_{T1}$ or $\alpha_{0\perp}'\hat{\upsilon} = \alpha_{0\perp}'(\hat{\Upsilon}^1 - \Upsilon^1(\hat{\pi}, \gamma_0))N_{T1}$, using Slutsky's Theorem.

3.5 Asymptotic distribution of the parameters in the extended model

We now apply Theorem 6 to derive the asymptotic distribution of the MLEs of the parameters in the extended core model, $\mathcal{H}_r^{ext*}$. By Theorem 4, these are the same as those of the MLEs of the same parameters in the full extended model, $\mathcal{H}_r^{ext}$.

Theorem 7. *Suppose Assumptions 1, 3, and 4 are satisfied and that the data is generated by (1) with* $\lambda = \lambda_0$*. We consider the statistical model* $\mathcal{H}_r^{ext*}$ *and the parameters* $\xi = (\pi, \rho, \Upsilon^1)$*. Then:*

(i) The asymptotic distribution of $T^{1/2}(\hat{\alpha} - \alpha_0, \hat{\Gamma}_1 - \Gamma_{1,0}, \ldots, \hat{\Gamma}_{k-1} - \Gamma_{k-1,0})$ *is given in (41) in Theorem 6, and the asymptotic distribution of* $T\bar{\beta}_{0\perp}'(\hat{\beta} - \beta_0)$ *follows from (42).*

(ii) For $\rho' = (\rho_1', \ldots, \rho_q')$*,* $\zeta = (\zeta_1', \zeta_2')'$ *defined in (37),* $\zeta_2' = (\zeta_{2,1}, \ldots, \zeta_{2,q})$*, and* $1 \le v \le q$*, the expansions*

$$\hat{\rho}_v' - \rho_{v0}' = -(\hat{\beta} - \beta_0)'\gamma_{0v}^0 - \hat{\beta}'(\hat{\gamma}_v^0 - \gamma_{0v}^0) = -(T^{1/2}\hat{\zeta}_1'\beta_{0\perp}'\gamma_{0v}^0)T^{-1} + (T^{1/2}\hat{\zeta}_{2,v})M_{T0v}^{-1}T^{-1/2} \tag{44}$$

shows that the asymptotic distribution of $\hat{\rho} - \rho_0$*, suitably normalized, follows from (42), and is mixed Gaussian.*

(iii) For $1 \le v \le q$ *and* $1 \le i \le m_v$*, the asymptotic distribution of* $\hat{\Upsilon}_{iv} - \Upsilon_{iv,0} = \hat{\Upsilon}_{iv} - \Upsilon_{iv}(\pi_0, \gamma_0)$ *consists of two terms,*

$$\hat{\Upsilon}_{iv} - \Upsilon_{iv,0} = (T^{1/2}\hat{v}_{iv})T^{-1/2}M_{Tiv}^{-1} + (T^{1/2}\Upsilon_{iv}(\hat{\pi} - \pi_0, \gamma_0))T^{-1/2}, \tag{45}$$

see (37). The asymptotic distribution of $T^{1/2}\hat{v}_{iv}$ *is given in (43), and, replacing* $\hat{\beta}$ *by* β_0*, the asymptotic distribution of* $T^{1/2}\Upsilon_{iv}(\hat{\pi} - \pi_0, \gamma_0)$ *depends only on that of* $\hat{\alpha}^*$*, which is Gaussian and given in (41), and the two terms on the right-hand side of (45) are asymptotically independent.*

The asymptotic distributions in parts (ii) and (iii) of Theorem 7 both depend only on the largest term. That is, in part (ii) the asymptotic distribution depends on the relation between the normalizations T^{-1} and $M_{T0v}^{-1}T^{-1/2}$ as well as the value of the parameter $\beta_{0\perp}'\gamma_{0v}^0$, whereas in part (iii) it depends on the relation between $T^{-1/2}M_{Tiv}^{-1}$ and $T^{-1/2}$. Haldrup (1996) encounters a similar problem of different limit behaviour of estimators in the context of a Dickey-Fuller regression with a slope coefficient.

3.6 Asymptotic distribution of the estimators in the additive model

Under Assumption 2, the extended and additive core models are reparametrizations of each other, see Theorem 4. Thus, in this case we have that $\Upsilon^1 = \Upsilon^1(\pi, \gamma)$, such that

$$\Upsilon^1(\hat{\pi}, \hat{\gamma}) - \Upsilon^1(\hat{\pi}, \gamma_0) = \Upsilon^1(\hat{\pi}, \hat{\gamma} - \gamma_0). \tag{46}$$

We use this fact, together with the simple one-to-one relation between the parameters in the extended core model, ξ, and the parameters in Theorem 6, and the one-to-one relation between ξ and $\lambda(\xi)$ in Theorem 1 to derive the asymptotic distribution of the maximum likelihood estimator for the additive model, which we here denote $\hat{\lambda} = \lambda(\hat{\xi})$, from the results in Theorem 6.

Theorem 8. *Suppose Assumptions 1–4 are satisfied and that the data is generated by (1) with* $\lambda = \lambda_0$*. We consider the statistical model* $\mathcal{H}_r^{add*}$ *and the parameters* $\lambda = (\pi, \gamma)$*. Then:*

(i) The asymptotic distributions of $T^{1/2}(\hat{\alpha} - \alpha_0, \hat{\Gamma}_1 - \Gamma_{1,0}, \ldots, \hat{\Gamma}_{k-1} - \Gamma_{k-1,0})$ *and* $(T(\hat{\beta} - \beta_0)'\bar{\beta}_{0\perp}, -T^{1/2}\beta_0'(\hat{\gamma}^0 - \gamma_0^0)N_{T0})$ *are given in (41) and (42) in Theorem 6.*

(ii) For $1 \le v \le q$ and $0 \le i \le m_v$, the asymptotic distributions of $T^{1/2}M_{Tiv}(\hat{\gamma}_v^i - \gamma_{0v}^i)$ are given in terms of those of $\hat{\upsilon}_{iv} = (\hat{\Upsilon}_{iv} - \Upsilon_{iv}(\hat{\pi}, \gamma_0))M_{Tiv}$, see (37) and (43), as follows,

$$T^{1/2}\beta_0'(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{Tiv} = -T^{1/2}\bar{\alpha}_0'\hat{\upsilon}_{iv} + o_P(1),\ 1 \le i \le m_v, \tag{47}$$

$$T^{1/2}\alpha_{0\perp}'\Gamma_0(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{T,i+1,v} = T^{1/2}\alpha_{0\perp}'\hat{\upsilon}_{i+1,v} + o_P(1),\ 0 \le i \le m_v - 1, \tag{48}$$

$$T^{1/2}\operatorname{vec}(\hat{\gamma}^{se} - \gamma_0^{se}) = (\sum_{i=0}^{k} M_s^{i\prime} \otimes \Phi_{i0})^{-1} T^{1/2}\operatorname{vec}(\hat{\upsilon}_{se}). \tag{49}$$

(iii) The distributions (47)–(49) are asymptotically independent of the distribution (41) given above.

The main result in Theorem 8 is that the simple condition in Assumption 2 of including enough differenced regressors in the additive model, implies that the additive and extended core models are reparametrizations of each other (Theorem 4). This permits relatively straightforward inference on the parameters of the additive model. Moreover, by Theorem 4, the asymptotic distributions of the MLEs of the parameters in the additive core model, $\mathcal{H}_r^{add*}$, given in Theorem 8, are the same as those for the same parameters in the full additive model, $\mathcal{H}_r^{add}$.

Recall that the distribution in (42) is mixed Gaussian (MG), so that asymptotic inference on β can be conducted using the χ^2-distribution also in the additive model. Also note that $T^{1/2}\alpha_{0\perp}'\hat{\upsilon}_{i+1,v}$ is asymptotically Gaussian, see (43), but $T^{1/2}\bar{\alpha}_0'\hat{\upsilon}_{iv}$ is neither asymptotically Gaussian nor mixed Gaussian. Thus, to use Theorem 8, for example, to test hypotheses on the parameter γ_v^0, the parameter needs to be divided into two components for which the estimators have different convergence rates as in (42) and (48). This is discussed in detail in Section 4.

3.7 Asymptotic distributions when $m > n$

If the condition that $\max_{1 \le v \le q} m_v \le n$ in Assumption 2 is violated, inference in the additive model becomes much more involved. To simplify the discussion, we consider the additive (core) model for a univariate regressor U_t with order m, lag length $k = 1$, and $n = 0$, that is $\gamma Z_t = \gamma U_t$. The general case follows similarly, but with more complicated notation.

We now consider the asymptotic distributions when $m > n$, i.e., when Assumption 2 is violated. The asymptotic theory for the extended model in Theorem 7 covers the case of $m > n$, but the theory for the additive model in Theorem 8 does not. In the next theorem, we compare inference in the two models in a simple case when $m > n$.

Theorem 9. *Suppose Assumptions 1, 3, and 4 are satisfied, but Assumption 2 is violated, and that the data is generated by (1) with $\lambda = \lambda_0$. We consider the statistical model $\mathcal{H}_r^{add*}$ in the special case $n = 0$, $k = 1$, and $m = 1$ and with the parameters $\lambda = (\pi, \gamma)$. Then:*

(i) The asymptotic distribution of $T^{1/2}(\hat{\alpha} - \alpha_0)$ in (41) continues to hold.

(ii) The asymptotic distribution of

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

or any linear combination of it, is not mixed Gaussian.

(iii) The asymptotic distribution $\alpha_{0\perp}'(\hat{\gamma} - \gamma_0)N_{T1}$ *is neither asymptotically Gaussian nor mixed Gaussian and the same holds for any linear combination of it.*

(iv) 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 $\hat{\alpha}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$ 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 $\alpha_\perp'\gamma$. Note that for $k = 1$ we have $\Gamma = I_p$, so that $\alpha_\perp'\beta_\perp$ has full rank. 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, and the cost of 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 when Assumption 2 is violated could possibly be solved by an application of the bootstrap along the lines of Cavaliere, Rahbek, and Taylor (2012) and Cavaliere, Nielsen, and Rahbek (2015). However, enlarging the model to have $n \geq m$ is a simple device to achieve simple inference as illustrated in Examples 4, 5, and 6.

4 Hypothesis testing

We first give the asymptotic distribution of the test for cointegration rank and then discuss tests on coefficients of deterministic terms. In both cases we work under Assumption 2 so that the additive and extended models are reparametrizations of each other.

4.1 Test of cointegration rank

We consider the extended model (18) for $r = p$, and regressors with bounded information removed,

$$\mathcal{H}_p^{ext*} : \Delta X_t = \Pi X_{t-1} + \Upsilon^0 Z_{0t} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon^1 Z_{1t} + \varepsilon_t. \quad (50)$$

The likelihood ratio test for rank r or $\Pi = \alpha\beta'$, where α and β are $p \times r$ matrices, is denoted $LR(\mathcal{H}_r^{ext*}|\mathcal{H}_p^{ext*})$. By Theorem 4, the asymptotic distribution of $LR(\mathcal{H}_r^{ext*}|\mathcal{H}_p^{ext*})$ is the same as that of $LR(\mathcal{H}_r^{ext}|\mathcal{H}_p^{ext})$ in the full model. 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 10. *Under Assumptions 1– 4, the asymptotic distribution of the test of cointegrating rank in either the extended core model, $\mathcal{H}_r^{ext*}$, or in the additive core model, $\mathcal{H}_r^{add*}$, is given by*

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

where $\varepsilon_{\alpha_\perp,t} = (\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1/2}\alpha_{0\perp}'\varepsilon_t$ is i.i.d. $(0, I_{p-r})$. By Theorem 4, the statistics $LR(\mathcal{H}_r^{ext}|\mathcal{H}_p^{ext})$ and $LR(\mathcal{H}_r^{add}|\mathcal{H}_p^{add})$ have the same asymptotic distribution as $LR(\mathcal{H}_r^{ext*}|\mathcal{H}_p^{ext*})$.

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 innovation formulation (6), this is not the case, see for example the analysis of the model with an unrestricted constant term in Johansen (1996, Chapter 13.5).

4.2 Tests of hypotheses on deterministic terms

We consider inference on the coefficients γ_v^i, $0 \leq i \leq m_v \leq n$, in the additive model under Assumption 2. It follows from Theorems 6 and 8 that the limit distribution of $\hat{\gamma}_v^i - \gamma_{v0}^i$ 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 $\alpha_\perp'\Gamma\gamma_v^i = 0$ assuming $\beta'\gamma_v^i = 0$ (since the matrix $(\beta, \Gamma'\alpha_\perp)$ has full rank under Assumption 1, see Theorem 1). It appears natural first to investigate if γ_v^0, i.e. 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 $\gamma_v^0 = 0$, etc. Thus we can apply the asymptotic distributions in Theorems 7 and 8 as in Theorem 11 to test recursively that $\gamma_v^i = 0$, provided we assume that $\gamma_v^j = 0$, $0 \leq j < i$.

Under Assumption 2 the additive and extended core models are reparametrizations (Theorem 4). Therefore, estimation of the unrestricted model can be performed by reduced rank regression of ΔX_t on $(X_{t-1}', Z_{0t}')'$ corrected for lagged ΔX_t and the regressors Z_{1t}, where $Z_{0t} = U_t$, see (25) and Section 2.6. Under the hypothesis $\beta'\gamma_v^0 = 0$, the parameters can be estimated in the same way by reduced rank regression, but removing U_{vt} from Z_{0t}. Finally, if both $\beta'\gamma_v^0 = 0$ and $\alpha_\perp'\Gamma\gamma_v^0 = 0$, so that $\gamma_v^0 = 0$, the estimation can also be performed by reduced rank regression, but replacing U_{vt} with ΔU_{vt} in Z_{0t} and removing ΔU_{vt} from Z_{1t}.

Theorem 11. *Let Assumptions 1–4 be satisfied. Then:*

- *(i) In the additive core model, $\mathcal{H}_r^{add*}$, the likelihood ratio test for the hypothesis $\rho_v' = -\beta'\gamma_v^0 = 0$ is asymptotically $\chi^2(r)$-distributed.*
- *(ii) In the additive core model, $\mathcal{H}_r^{add*}$, with $m_v \geq 1$, the likelihood ratio test for the hypothesis $\alpha_\perp'\Gamma\gamma_v^0 = 0$, given that $\rho_v' = 0$, is asymptotically $\chi^2(p-r)$-distributed.*
- *(iii) In the additive core model, $\mathcal{H}_r^{add*}$, with $m_v \geq 1$, the likelihood ratio test for the joint hypothesis, $\gamma_v^0 = 0$, is asymptotically $\chi^2(p)$-distributed.*
- *(iv) By Theorem 4, the same results hold in the additive model, $\mathcal{H}_r^{add}$.*

5 Conclusions

We define the CVAR model with additive deterministic terms and derive the corresponding innovation 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 innovation 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 2, 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 quite strange in most circumstances. 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 4. However, the recommendation is to include the first difference anyway, even if it is believed to be zero, because including it leads to simple inference, see Examples 5 and 6.

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

A Appendix: proofs of results

A.1 Proof of Theorem 1

The two results in (15) follow from (14) when multiplying by $\bar{\alpha}'$ and $\alpha_\perp'$, respectively, using $\Phi_0 = -\alpha\beta'$ and $\Phi_1 = \alpha\beta' + \Gamma$. It follows from

$$(\beta, \Gamma'\alpha_\perp)'(\beta, \beta_\perp) = \begin{pmatrix} \beta'\beta & 0 \\ \alpha_\perp'\Gamma\beta & \alpha_\perp'\Gamma\beta_\perp \end{pmatrix}, \tag{51}$$

and Assumption 1 that $\alpha_\perp'\Gamma\beta_\perp$ has full rank, and hence that also $(\beta, \Gamma'\alpha_\perp)$ has full rank, so that the relations in (15) can be solved for γ^i. 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

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

Now $\omega_j = 1 - e^{2\pi \mathrm{i} j/s}$, $j = 1, \ldots, s-1$, see Example 3, such that $\Pi(1 - \omega_j) = \Pi(e^{2\pi \mathrm{i} j/s})$, which by Assumption 1 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

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

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 2

The result $\Delta f_i(t) = f_{i-1}(t)$ follows from the identity

$$\begin{aligned}(t+i)\cdots(t+1) - (t-1+i)\cdots t &= (t+i)\left((t-1+i)\cdots(t+1)\right) - \left((t-1+i)\cdots(t+1)\right)t \\ &= \left((t-1+i)\cdots(t+1)\right)(t+i-t) = \left((t-1+i)\cdots(t+1)\right)i\end{aligned}$$

after division by $i!$.

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

$$\begin{aligned}\Pi(L)\gamma Z_t &= \sum_{i=0}^{k} \Phi_i \Delta^i \sum_{j=0}^{m} \gamma^j f_{m-j}(t) = \sum_{i=0}^{k}\sum_{j=0}^{m} \Phi_i \gamma^j f_{m-j-i}(t) \\ &= \sum_{s=0}^{m} \Upsilon_s f_{m-s}(t) = \alpha\rho' f_m(t) + \sum_{s=1}^{m} \Upsilon_s f_{m-s}(t),\end{aligned}$$

where the penultimate equality follows because $\Delta^j f_m(t) = 0$ for $j > m$.

A.3 Proof of Theorem 3

The proof follows from Theorem 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 Lemma 1

For $T_1 \leq T$, we decompose the numerator as

$$\sum_{t=1}^{T} Z_{1t} Z_{2t} = \sum_{t=1}^{T_1} Z_{1t} Z_{2t} + \sum_{t=T_1+1}^{T} Z_{1t} Z_{2t},$$

and evaluate the second term using the Cauchy-Schwarz inequality,

$$|\sum_{t=T_1+1}^{T} Z_{1t} Z_{2t}| \leq (\sum_{t=T_1+1}^{T} Z_{1t}^2)^{1/2} (\sum_{t=T_1+1}^{T} Z_{2t}^2)^{1/2} \leq (\sum_{t=1}^{T} Z_{1t}^2)^{1/2} (\sum_{t=T_1+1}^{\infty} Z_{2t}^2)^{1/2}.$$

Because $\sum_{t=1}^{T} Z_{2t}^2 \leq c$, it follows that, for all $\epsilon > 0$ and T_1 sufficiently large, we have

$$|\sum_{t=T_1+1}^{T} Z_{1t} Z_{2t}| (\sum_{t=1}^{T} Z_{1t}^2)^{-1/2} \leq (\sum_{t=T_1+1}^{\infty} Z_{2t}^2)^{1/2} \leq \epsilon/2.$$

Finally, because $\sum_{t=1}^{T} Z_{1t}^2 \overset{\mathrm{P}}{\to} \infty$, we can choose T so large that, for any (fixed) T_1 and any $\delta > 0$, we have

$$|\sum_{t=1}^{T_1} Z_{1t} Z_{2t}| (\sum_{t=1}^{T} Z_{1t}^2)^{-1/2} \leq \epsilon/2$$

with probability greater than $1 - \delta$, which completes the proof.

A.5 Proof of Theorem 4

Lemma 1 shows that, for both the additive and the extended models, the limit of the information matrix normalized by its diagonal elements is asymptotically block diagonal. One block corresponds to the coefficients of the regressors with bounded information and another corresponds to the parameters in the core model. This implies that the asymptotic distributions of the MLEs in the core models are the same as the asymptotic distributions of the MLEs of the same parameters in the full models.

Under Assumptions 1 and 2, the core models, $\mathcal{H}_r^{ext*}$ and $\mathcal{H}_r^{add*}$, have the same number of parameters. For the core model $\mathcal{H}_r^{ext*}$, see (27) the parameters are, apart from π and Υ_{se}, the $rq + \sum_{v=1}^{q} m_v$ parameters collected in

$$(\rho, \Upsilon_{i1}, \ldots, \Upsilon_{i,m_v}, 1 \leq v \leq q). \tag{52}$$

For the core model $\mathcal{H}_r^{add*}$, the parameters are, see (11), apart from π and Υ_{se}, the $pq + \sum_{i=1}^{q}(p(m_v - 1) + r)$ parameters collected in

$$(\gamma^0, \gamma_v^1, \ldots, \gamma_v^{m_v-1}, \beta'\gamma_v^{m_v}, 1 \leq v \leq q). \tag{53}$$

Theorem 1 shows how (52) can be recovered from (53) and vice versa. Thus, the core models are reparametrizations of each other.

A.6 Proof of Theorem 5

We can express both the additive model (8) and the extended model (18) 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 (18) 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

$$\Delta X_t = \alpha(\beta_0' X_{t-1}) + \phi(\beta_{0\perp}' X_{t-1}) + \psi Z_{0t} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon^1 Z_{1t} + \Upsilon^2 Z_{2t} + \varepsilon_t \tag{54}$$

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 3 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 (54). 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 Theorems 3 and 4.

A.7 Proof of Theorem 6

The likelihood is given in (31) with $\varepsilon_t(\theta)$ replaced by the residuals $\varepsilon_t(\alpha^*, \zeta, \upsilon)$, see 40).

Derivatives of ε_t: At the true values, $(\alpha^*, \zeta, \upsilon) = (\alpha_0^*, \zeta_0, \upsilon_0) = (0, 0, 0)$, we find the derivatives

$$\begin{aligned} D_{\alpha^*}\varepsilon_t(\alpha_0^*, \zeta_0, \upsilon_0; d\alpha^*) &= -(d\alpha^*)Y_t^*, \\ D_{\zeta}\varepsilon_t(\alpha_0^*, \zeta_0, \upsilon_0; d\zeta) &= -\alpha_0(d\zeta)'G_{Tt}, \\ D_{\upsilon}\varepsilon_t(\alpha_0^*, \zeta_0, \upsilon_0; d\upsilon) &= -(d\upsilon)Z_{1Tt}. \end{aligned}$$

Score and information: We denote the score function with respect to α^*, for example, in the direction $d\alpha^*$ as $S_{\alpha^*} = D_{\alpha^*} \log L(\alpha_0^*, \zeta_0, \upsilon_0; d\alpha^*)$. The information with respect to α^* and ζ, for example, is similarly denoted by $I_{\alpha^*\zeta} = D^2_{\alpha^*\zeta} \log L(\alpha_0^*, \zeta_0, \upsilon_0; d\alpha^*, d\zeta)$. Then the scores are

$$T^{-1/2} S_{\alpha^*} = -\operatorname{tr}\{\Omega_0^{-1}(d\alpha^*) T^{1/2} \langle Y^*, \varepsilon\rangle_T\} \xrightarrow{\mathrm{D}} -\operatorname{tr}\{\Omega_0^{-1}(d\alpha^*) \langle Y^*, \varepsilon\rangle\}, \tag{55}$$

$$T^{-1/2} S_{\zeta} = -\operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)' T^{1/2} \langle G_T, \varepsilon\rangle_T\} \xrightarrow{\mathrm{D}} -\operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)' \langle G, \varepsilon\rangle\}, \tag{56}$$

$$T^{-1/2} S_{\upsilon} = -\operatorname{tr}\{\Omega_0^{-1}(d\upsilon) T^{1/2} \langle Z_{1T}, \varepsilon\rangle_T\} \xrightarrow{\mathrm{D}} -\operatorname{tr}\{\Omega_0^{-1}(d\upsilon) \langle Z_1, \varepsilon\rangle\}, \tag{57}$$

and the diagonal elements of the information are

$$T^{-1} I_{\alpha^*\alpha^*} = \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*) \langle Y^*, Y^*\rangle_T (d\alpha^*)'\} + o_P(1) \xrightarrow{\mathrm{P}} \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\Sigma_{stat}(d\alpha^*)'\}, \tag{58}$$

$$T^{-1} I_{\zeta\zeta} = \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)' \langle G_T, G_T\rangle_T (d\zeta)\alpha_0'\} + o_P(1) \xrightarrow{\mathrm{D}} \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)' \langle G, G\rangle (d\zeta)\alpha_0'\}, \tag{59}$$

$$T^{-1} I_{\upsilon\upsilon} = \operatorname{tr}\{\Omega_0^{-1}(d\upsilon) \langle Z_{1T}, Z_{1T}\rangle_T (d\upsilon)'\} + o_P(1) \xrightarrow{\mathrm{D}} \operatorname{tr}\{\Omega_0^{-1}(d\upsilon) \langle Z_1, Z_1\rangle (d\upsilon)'\}.$$

There is one non-zero off-diagonal element,

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

and the following are asymptotically negligible,

$$T^{-1} I_{\alpha^*\zeta} = \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*) \langle Y^*, G_T\rangle_T (d\zeta)\alpha_0'\} + o_P(1) \xrightarrow{\mathrm{P}} 0, \tag{60}$$

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

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).

The asymptotic distribution of $T^{1/2}\hat{\alpha}^*$: By the usual Taylor expansion of the likelihood equations, we find that the equation for the asymptotic distribution of $T^{1/2}\hat{\alpha}^*$ is 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 (41).

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

$$\langle G_T, G_T\rangle_T (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1}\alpha_0 + \langle G_T, Z_{1T}\rangle_T (T^{1/2}\hat{\upsilon})'\Omega_0^{-1}\alpha_0 = -T^{1/2} \langle G_T, \varepsilon\rangle_T \Omega_0^{-1}\alpha_0 + o_P(1), \tag{61}$$

$$\langle Z_{1T}, G_T\rangle_T (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1} + \langle Z_{1T}, Z_{1T}\rangle_T (T^{1/2}\hat{\upsilon})'\Omega_0^{-1} = -T^{1/2} \langle Z_{1T}, \varepsilon\rangle_T \Omega_0^{-1} + o_P(1). \tag{62}$$

Pre-multiplying (62) by $\langle G_T, Z_{1T}\rangle_T \langle Z_{1T}, Z_{1T}\rangle_T^{-1}$, post-multiplying by α_0, and subtracting the result from (61) 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 (42) and inserted into (62) gives (43).

A.8 Proof of Theorem 7

First note that $\mathcal{H}_r^{ext*}$ with the parameters ξ is a simple one-to-one reparametrization of $\mathcal{H}_r^{ext*}$ with the parameters $(\alpha^*, \zeta, \upsilon)$. Thus, these parametrizations have the same likelihood and the MLEs of ξ can easily be found from those of $(\alpha^*, \zeta, \upsilon)$ and vice versa.

Proof of (i): The result for the dynamic parameters $\pi = (\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1})$ follows directly from (41) and (42) by the definitions of α^* and ζ, see (37) and (38).

Proof of (ii): Using $\rho_v' = -\beta'\gamma_v^0$ and the definition of $\zeta = (\zeta_1', \zeta_2')'$ in (37), we find the expansion (44). This expansion shows that the limit is a linear combination of the elements of the limit of $\hat{\zeta}$, and hence mixed Gaussian, see (42).

Proof of (iii): The expansion (45) follows directly from the definition of $\hat{\upsilon}_{iv}$ in (37). Clearly the limit of the first term, $T^{1/2}\hat{\upsilon}_{iv}$, is given in (43). Because $\hat{\beta}$ converges at rate T (super consistency), see (42), we can replace $\hat{\beta}$ by β_0 in the second term, $T^{1/2}\Upsilon_{iv}(\hat{\pi} - \pi_0, \gamma_0)$, without changing the asymptotic properties. Hence, $T^{1/2}\Upsilon_{iv}(\hat{\pi} - \pi_0, \gamma_0)$ becomes a linear combination of $T^{1/2}\hat{\alpha}^*$, see (38), and is therefore asymptotically Gaussian. Asymptotic independence of the two terms follows because the limits in (42) and (43) are independent, see Theorem 6.

A.9 Proof of Theorem 8

Under Assumption 2, the extended and additive core models are reparametrizations of each other, see Theorem 4. Moreover, the simple one-to-one relation between the parameters in the extended core model, ξ, and the parameters in Theorem 6 then show that in fact the additive core model and the extended core model parametrized by $(\alpha^*, \zeta, \upsilon)$ are reparametrizations. Hence, they have the same likelihood and the MLEs of the parameters of the additive model can be found directly from the MLEs in Theorem 6 using Theorem 1.

Proof of (i): First, $\beta_0'(\hat{\gamma}^0 - \gamma_0^0)N_{T0} = \hat{\beta}'(\hat{\gamma}^0 - \gamma_0^0)N_{T0} + o_P(1)$ by Slutsky's Theorem and consistency of $\hat{\beta}$. Under Assumption 2 it follows from Theorem 4 that $\hat{\beta}'(\hat{\gamma}^0 - \gamma_0^0)N_{T0} = -(\hat{\rho}' + \hat{\beta}'\gamma_0^0)N_{T0} = \hat{\zeta}_2'$, where $\zeta = (\zeta_1', \zeta_2')'$, see (37). The results for the dynamic parameters $\pi = (\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1})$ and for $\beta_0'(\hat{\gamma}^0 - \gamma_0^0)N_{T0}$ then follow directly from (41) and (42) by the definitions of α^* and ζ, see (37) and (38).

Proof of (ii): Again, by Slutsky's Theorem and consistency of $\hat{\beta}$ it follows that $T^{1/2}\beta_0'(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{Tiv} = T^{1/2}\hat{\beta}'(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{Tiv} + o_P(1)$ and $T^{1/2}\alpha_{0\perp}'\Gamma_0(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{T,i+1,v} = T^{1/2}\hat{\alpha}_\perp'\hat{\Gamma}(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{T,i+1,v} + o_P(1)$. Next, we note that, under Assumption 2, $\hat{\upsilon}_{iv} = (\hat{\Upsilon}_{iv} - \Upsilon_{iv}(\hat{\pi}, \gamma_0)) = \Upsilon_{iv}(\hat{\pi}, \hat{\gamma} - \gamma_0)$ by (46) and Theorem 4. By Theorem 1, applied to $(\hat{\pi}, \gamma_0)$ and $(\hat{\pi}, \hat{\gamma})$, we then find that $\hat{\beta}'(\hat{\gamma}_v^i - \gamma_{v0}^i)$ for $i = 1, \ldots, m_v$ and $\hat{\alpha}_\perp'\hat{\Gamma}(\hat{\gamma}_v^i - \gamma_{v0}^i)$ for $i = 0, \ldots, m_v - 1$ can be expressed in terms of $\hat{\upsilon}_{iv}$ as

$$
\begin{aligned}
T^{1/2}\hat{\beta}'(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{Tiv} &= -T^{1/2}\hat{\bar{\alpha}}_0'\Upsilon_{iv}(\hat{\pi}, \hat{\gamma} - \gamma_0)M_{Tiv} + T^{1/2}\hat{\bar{\alpha}}' \sum_{j=1}^{\min\{i,k\}} \hat{\Phi}_j(\hat{\gamma}_v^{i-j} - \gamma_{v0}^{i-j})M_{Tiv} \\
&= -T^{1/2}\hat{\bar{\alpha}}'\Upsilon_{iv}(\hat{\pi}, \hat{\gamma} - \gamma_0)M_{Tiv} + o_P(1) \\
&= -T^{1/2}\hat{\bar{\alpha}}'\hat{\upsilon}_{iv} + o_P(1) = -T^{1/2}\bar{\alpha}_0'\hat{\upsilon}_{iv} + o_P(1)
\end{aligned}
$$

using that $M_{T,i-j,v}^{-1}M_{Tiv} \to 0$ for $j = 1, \ldots, \min\{i, k\}$ (by Assumption 3) and using Slutsky's

Theorem again for the final equality. Similarly, for $i = 0, \dots, m_v - 1$,

$$\begin{aligned}T^{1/2}\hat{\alpha}_\perp'\hat{\Gamma}(\hat{\gamma}_v^i - \gamma_{v0}^i)M_{T,i+1,v} =& T^{1/2}\hat{\alpha}_\perp'\Upsilon_{i+1,v}(\hat{\pi}, \hat{\gamma} - \gamma_0)M_{T,i+1,v} \\ &- T^{1/2}\hat{\alpha}_\perp' \sum_{j=2}^{\min\{i+1,k\}} \hat{\Phi}_j(\hat{\gamma}_v^{i+1-j} - \gamma_{v0}^{i+1-j})M_{T,i+1,v} \\ =& T^{1/2}\hat{\alpha}_\perp'\Upsilon_{i+1,v}(\hat{\pi}, \hat{\gamma} - \gamma_0)M_{T,i+1,v} + o_P(1) \\ =& T^{1/2}\hat{\alpha}_\perp'\hat{\upsilon}_{i+1,v} + o_P(1) = T^{1/2}\alpha_{0\perp}'\hat{\upsilon}_{i+1,v} + o_P(1)\end{aligned}$$

because $M_{T,i-j+1,v}^{-1}M_{T,i+1,v} \to 0$ for $j = 2, \dots, \min\{i+1, k, m_v\}$. This proves (47) and (48). Finally, (49) follows directly from vectorization of (14) and noting that $\sum_{i=0}^k M_s^{i\prime} \otimes \Phi_{0i}$ is invertible by Theorem 3.

Proof of (iii): That (47)–(49) are asymptotically independent of (41), is shown in Theorem (6).

A.10 Proof of Theorem 9

Proof of (i): Because $k = 1$ and $n = 0$ we have $U_t = Z_t$, $\gamma^0 = \gamma$, and

$$\varepsilon_t(\lambda) = \Delta X_t - \alpha(\beta' X_{t-1} - \beta'\gamma Z_{t-1}) - \gamma\Delta Z_t.$$

Furthermore, $\alpha_{0\perp}'\Gamma_0\beta_{0\perp} = \alpha_{0\perp}'\beta_{0\perp}$ and $\gamma - \gamma_0$ can be decomposed as

$$\gamma - \gamma_0 = \alpha(\beta'\alpha)^{-1}\beta'(\gamma - \gamma_0) + \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'(\gamma - \gamma_0). \tag{63}$$

We define $\zeta_2' = -\beta'(\gamma - \gamma_0)M_{T0}$ and $\phi' = T^{1/2}(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'(\gamma - \gamma_0)M_{T1}$ as well as $\alpha^* = \alpha - \alpha_0, Y_t^* = \beta_0' Y_{t-1}$, see (37) and (38). The corresponding likelihood is then based on

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

see (40). We note that ζ_2 appears in two places, but $M_{T0}^{-1}M_{T1} \to 0$, and therefore the term $\zeta_2' M_{T0}^{-1}M_{T1}Z_{1Tt}$ disappears in the asymptotic analysis of the score and information, because it is dominated by the term $\alpha\zeta' G_{Tt}$.

Mimicking the analysis in the proof of Theorem 6, we find at the true values, $\alpha_0^* = 0, \zeta_0 = 0, \phi_0 = 0$, the derivatives

$$\begin{aligned}&D_{\alpha^*}\varepsilon_t(\alpha_0^*, \zeta_0, \phi_0; d\alpha^*) = -(d\alpha^*)Y_t^*, \\ &D_\zeta\varepsilon_t(\alpha_0^*, \zeta_0, \phi_0; d\zeta) = -\alpha_0(d\zeta)' G_{Tt} - \alpha_0(\beta_0'\alpha_0)^{-1}(d\zeta_2)' M_{T0}^{-1}M_{T1}Z_{1Tt} = -\alpha_0(d\zeta)' G_{Tt} + o(1), \\ &D_\phi\varepsilon_t(\alpha_0^*, \zeta_0, \phi_0; d\phi) = -\beta_{0\perp}(d\phi)' Z_{1Tt}.\end{aligned}$$

The limits of the scores for α^* and ζ are therefore given in (55) and (56), and for ϕ we find

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

The limits of the information matrix blocks $I_{\alpha^*\alpha^*}$, $I_{\zeta\zeta}$, and $I_{\alpha^*\zeta}$ are given in (58), (59), and (60), respectively, and for ϕ we find

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

Thus, the only difference compared with the extended model is the factor $\beta_{0\perp}$, which comes from estimating $\phi = (\alpha_{0\perp}'\beta_{0\perp})^{-1}\alpha_{0\perp}'(\gamma - \gamma_0)$ as the coefficient to ΔZ_t. It is seen that the limit information is block-diagonal corresponding to α^* and (ζ, ϕ), such that the asymptotic distribution of $T^{1/2}\hat{\alpha}^* = T^{1/2}(\hat{\alpha} - \alpha_0)$ is as given in (41) in Theorem 6.

Proof of (ii): We find the equations for determining the limit distribution of the MLE $T^{1/2}(\hat{\zeta}, \hat{\phi})$, but compared with (61) and (62) there is now an extra factor $\beta_{0\perp}$ in (65),

$$\langle G, G\rangle (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1}\alpha_0 + \langle G, Z_1\rangle (T^{1/2}\hat{\phi})\beta_{0\perp}'\Omega_0^{-1}\alpha_0 \overset{\mathrm{D}}{\rightarrow} - \langle G, \varepsilon\rangle \Omega_0^{-1}\alpha_0, \tag{64}$$

$$\langle Z_1, G\rangle (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1}\beta_{0\perp} + \langle Z_1, Z_1\rangle (T^{1/2}\hat{\phi})\beta_{0\perp}'\Omega_0^{-1}\beta_{0\perp} \overset{\mathrm{D}}{\rightarrow} - \langle Z_1, \varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}. \tag{65}$$

Eliminating $T^{1/2}\hat{\phi}$ from (64), 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

$$E(\langle Z_1, \varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}|\alpha_{0\perp}'W_\varepsilon) = \langle Z_1, \varepsilon\rangle \alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'\beta_{0\perp} \neq 0. \tag{66}$$

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

Proof of (iii): If we eliminate $T^{1/2}\hat{\zeta}$ from the equations (64) and (65), 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 this distribution has mean $-E(\langle Z_1, \varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}|\alpha_{0\perp}'W_\varepsilon)$, see (66), and the limit distribution of $T^{1/2}\hat{\phi}$ is neither Gaussian nor mixed Gaussian, and the same holds for any linear combination.

Proof of (iv): From (61) and (62) we find the equations to determine the limit distribution of the MLE in the extended model $T^{1/2}(\hat{\zeta}, \hat{\upsilon})$,

$$\langle G, G\rangle (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1}\alpha_0 + \langle G, Z_1\rangle (T^{1/2}\hat{\upsilon})\Omega_0^{-1}\alpha_0 \overset{\mathrm{D}}{\rightarrow} - \langle G, \varepsilon\rangle \Omega_0^{-1}\alpha_0,$$

$$\langle Z_1, G\rangle (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1} + \langle Z_1, Z_1\rangle (T^{1/2}\hat{\upsilon})\Omega_0^{-1} \overset{\mathrm{D}}{\rightarrow} - \langle Z_1, \varepsilon\rangle \Omega_0^{-1},$$

and from these we find that the limit of the information matrix for (ζ, υ) in the extended model is,

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

say. In the additive model, the limit of the information matrix for (ζ, ϕ) is, see (64) and (65),

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

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

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

where the inequality means that the difference is positive definite, almost surely. It follows that the asymptotic conditional variance of $\hat{\zeta}$ in the additive model is larger than that of $\hat{\zeta}$ in the extended model because

$$AsVar(\hat{\zeta}_{add}) = (I^{add}_{\zeta\zeta} - I^{add}_{\zeta\phi}(I^{add}_{\phi\phi})^{-1}I^{add}_{\phi\zeta})^{-1} < (I^{ext}_{\zeta\zeta} - I^{ext}_{\zeta\upsilon}(I^{ext}_{\upsilon\upsilon})^{-1}I^{ext}_{\upsilon\zeta})^{-1} = AsVar(\hat{\zeta}_{ext}),$$

almost surely.

A.11 Proof of Theorem 10

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

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

and define the auxiliary hypothesis

$$\mathcal{H} = \{\Pi\bar{\beta}_{0\perp} = 0 \text{ and } \beta_0'(\gamma^0 - \gamma^0_0) = 0\}.$$

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

$$\mathcal{H}^{ext*}_p \cap \mathcal{H} = \mathcal{H}^{ext*}_r \cap \{\beta^* = \beta^*_0\}.$$

To facilitate the analysis of the test for rank, we introduce the extra hypothesis $\mathcal{H}$ in models $\mathcal{H}^{ext*}_p$ and $\mathcal{H}^{ext*}_r$, see Lawley (1956) for an early application of this idea or Johansen (2002, p. 1947) and Johansen and Nielsen (2012, p. 2699) for applications to the (fractional) CVAR model. We then find

$$LR(\mathcal{H}^{ext*}_r|\mathcal{H}^{ext*}_p) = \frac{\max_{\mathcal{H}^{ext*}_r} L_T(\xi)}{\max_{\mathcal{H}^{ext*}_p} L_T(\xi)} = \frac{\max_{\mathcal{H}^{ext*}_p \cap \mathcal{H}} L_T(\xi)}{\max_{\mathcal{H}^{ext*}_p} L_T(\xi)} \Big/ \frac{\max_{\mathcal{H}^{ext*}_r \cap \{\beta^*=\beta^*_0\}} L_T(\xi)}{\max_{\mathcal{H}^{ext*}_r} L_T(\xi)}.$$

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

$$LR(\mathcal{H}^{ext*}_r|\mathcal{H}^{ext*}_p) = \frac{LR(\mathcal{H}|\mathcal{H}^{ext*}_p)}{LR(\beta^* = \beta^*_0|\mathcal{H}^{ext*}_r)}. \tag{67}$$

Analysis of $LR(\mathcal{H}|\mathcal{H}^{ext*}_p)$: We apply the formulas (35) and (36) for $\mathcal{H}^{ext*}_p$, using $\alpha - \alpha_0 = (\Pi - \Pi_0)\bar{\beta}_0$, to find

$$\begin{aligned}
-\Pi(L)(\gamma - \gamma_0)Z_t &= \Pi(\gamma^0 - \gamma^0_0)N_{T0}Z_{0Tt} - \Upsilon^1(\pi, \gamma - \gamma_0)N_{T1}Z_{1Tt}, \\
(\Pi(L) - \Pi_0(L))Y_t &= -\alpha^*Y^*_t - T^{1/2}\Pi\bar{\beta}_{0\perp}T^{-1/2}\beta_{0\perp}'Y_{t-1},
\end{aligned}$$

where α^* and Y_t^* are given in (38). Defining

$$\zeta' = (T^{1/2}\Pi\bar{\beta}_{0\perp}, \Pi(\gamma^0 - \gamma_0^0)N_{T0}) \text{ and } G_{Tt} = \begin{pmatrix} T^{-1/2}\beta'_{0\perp}Y_{t-1} \\ Z_{0Tt} \end{pmatrix}, \tag{68}$$

we then obtain the residuals

$$\varepsilon_t(\xi) = -\alpha^* Y_t^* - \zeta' G_{Tt} - \Upsilon^1(\pi, \gamma - \gamma_0)N_{T1}Z_{1Tt} + \varepsilon_t,$$

see also (40). 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 $\langle\varepsilon,\varepsilon|Z_{1T}\rangle_T \xrightarrow{\text{P}} \Omega_0$ that

$$\begin{aligned}-2\log LR(\mathcal{H}|\mathcal{H}_p^{ext*}) &= \text{tr}\{\Omega_0^{-1}T^{1/2}\langle\varepsilon, G_T|Z_{1T}\rangle_T \langle G_T, G_T|Z_{1T}\rangle_T^{-1} T^{1/2}\langle G_T,\varepsilon|Z_{1T}\rangle_T\} + o_P(1) \\ &\xrightarrow{\text{D}} \text{tr}\{\Omega_0^{-1}\langle\varepsilon, G|Z_1\rangle \langle G, G|Z_1\rangle^{-1}\langle G,\varepsilon|Z_1\rangle\}.\end{aligned} \tag{69}$$

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^* is asymptotically independent of inference on (β, γ^0), we can assume that $\alpha^* = \alpha_0^* = 0$ for the asymptotic analysis of this statistic. We now find for $\beta = \beta_0 + \beta_{0\perp}\bar{\beta}'_{0\perp}(\beta - \beta_0)$ that

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

We define G_{Tt} as above, see (68), and define

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

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

$$\mathcal{H}_r^{ext*} : \varepsilon_t(\xi) = -\alpha_0\zeta' G_{Tt} - \Upsilon^1(\pi, \gamma-\gamma_0)N_{T1}Z_{1Tt} + \varepsilon_t.$$

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

$$\begin{aligned}\alpha'_{\Omega_0}\varepsilon_t(\xi) &= -\zeta' G_{Tt} - \alpha'_{\Omega_0}\Upsilon^1(\pi, \gamma-\gamma_0)N_{T1}Z_{1Tt} + \alpha'_{\Omega_0}\varepsilon_t, \\ \alpha'_{0\perp}\varepsilon_t(\xi) &= -\alpha'_{0\perp}\Upsilon^1(\pi, \gamma-\gamma_0)N_{T1}Z_{1Tt} + \alpha'_{0\perp}\varepsilon_t.\end{aligned}$$

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

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

and

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

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

$$\begin{aligned}
&-2\log LR(\beta^* = \beta_0^*|\mathcal{H}_r^{ext*}) \\
&\quad = \mathrm{tr}\{\alpha_0'\Omega_0^{-1}\alpha_0 T^{1/2}\left\langle \alpha_{\Omega_0}'\varepsilon, G_T|Z_{1T}\right\rangle_T \left\langle G_T, G_T|Z_{1T}\right\rangle_T^{-1} T^{1/2}\left\langle G_T, \alpha_{\Omega_0}'\varepsilon|Z_{1T}\right\rangle_T\} + o_P(1) \\
&\quad \xrightarrow{\mathrm{D}} \mathrm{tr}\{\Omega_0^{-1}\alpha_0(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\left\langle \varepsilon, G|Z_1\right\rangle \left\langle G, G|Z_1\right\rangle^{-1}\left\langle G, \varepsilon|Z_1\right\rangle\}. \qquad (70)
\end{aligned}$$

Analysis of $LR(\mathcal{H}_r^{ext*}|\mathcal{H}_p^{ext*})$: By (67), the test for rank r converges to the difference between (69) and (70), i.e.,

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

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.12 Proof of Theorem 11

Proof of (i): We find from Theorem 7(ii) that $\hat{\rho}_v' - \rho_{v0}'$, suitably normalized, is mixed Gaussian. Hence, the likelihood ratio test for $\beta'\gamma_v^0 = -\rho_v' = 0$ is a $\chi^2(r)$ test.

Proof of (ii): We test the hypothesis that $\alpha_\perp'\Gamma\gamma_v^0 = 0$ using the statistic $\hat{\alpha}_\perp'\hat{\Gamma}\hat{\gamma}_v^0$, assuming that $\beta_0'\gamma_{v0}^0 = -\rho_{v0}' = 0$. Under the null that also $\alpha_{0\perp}'\Gamma_0\gamma_{v0}^0 = 0$, we find that $\gamma_{v0}^0 = 0$ because $(\beta_0, \Gamma_0'\alpha_{0\perp})$ has full rank by Assumption 1, and it follows that $\hat{\alpha}_\perp'\hat{\Gamma}\hat{\gamma}_v^0 = \hat{\alpha}_\perp'\hat{\Gamma}(\hat{\gamma}_v^0 - \gamma_{0v}^0)$. From Slutsky's Theorem and consistency of $\hat{\pi}$ (Theorem 5), we then find

$$T^{1/2}\hat{\alpha}_\perp'\hat{\Gamma}(\hat{\gamma}_v^0 - \gamma_{v0}^0)M_{T1v} = T^{1/2}\alpha_{0\perp}'\Gamma_0(\hat{\gamma}_v^0 - \gamma_{v0}^0)M_{T1v} + o_P(1).$$

The asymptotic distribution of the first term on the right-hand side is found from (48) in Theorem 8 and (43) in Theorem 6,

$$T^{1/2}\alpha_{0\perp}'\Gamma_0(\hat{\gamma}_v^0 - \gamma_{v0}^0)M_{T1v} = T^{1/2}\alpha_{0\perp}'\hat{v}_{1v} + o_P(1) \xrightarrow{\mathrm{D}} -\alpha_{0\perp}'\left\langle \varepsilon, Z_1\right\rangle \left\langle Z_1, Z_1\right\rangle^{-1} A$$

for a suitable selection matrix A. This limit is Gaussian so that $T^{1/2}\hat{\alpha}_\perp'\hat{\Gamma}(\hat{\gamma}_v^0 - \gamma_{v0}^0)M_{T1v}$ is asymptotically Gaussian, and hence the likelihood ratio test for $\alpha_\perp'\Gamma\gamma_v^0 = 0$ is asymptotically $\chi^2(p-r)$.

Proof of (iii): Follows trivially from (i) and (ii).

Acknowledgements

We are grateful to two anonymous referees, Giuseppppe Cavaliere, Anders Rahbek, and 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. (2015). 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. (2000). Testing for the cointegrating rank of a VAR process with a time trend. *Journal of Econometrics* **95**, pp. 177–198.
16. Miller, K.S. (1968). *Linear Difference Equations*, Benjamin, New York.
17. Nielsen, H.B. (2004). Cointegration analysis in the presence of outliers. *Econometrics Journal* **7**, pp. 249–271.
18. Nielsen, H.B. (2007). UK money demand 1873-2011: a long time series analysis and event study. *Cliometrics* **1**, pp. 45–61.
19. Nielsen, B., and Rahbek, A. (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics* **62**, pp. 5–22.
20. 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.

21. 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.
22. Rahbek, A., and Mosconi, R. (1999). Cointegration rank inference with stationary regressors in VAR models. *Econometrics Journal* **2**, pp. 76–91.
23. Saikkonen, P., and Lütkepohl, H. (2000a). Testing for the cointegrating rank of a VAR process with an intercept. *Econometric Theory* **16**, pp. 373–406.
24. Saikkonen, P., and Lütkepohl, H. (2000b). Testing for the cointegrating rank of a VAR process with structural shifts. *Journal of Business & Economic Statistics* **18**, pp. 451–464.
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.