Vector Autoregresive Moving Average Identification for Macroeconomic Modeling: A New Methodology

D. S. Poskitt*

Department of Econometrics & Business Statistics, Monash University

Abstract

This paper develops a new methodology for identifying the structure of VARMA time series models. The analysis proceeds by examining the echelon canonical form and presents a fully automatic data driven approach to model specification using a new technique to determine the Kronecker invariants. A novel feature of the inferential procedures developed here is that they work in terms of a canonical scalar ARMAX representation in which the exogenous regressors are given by predetermined contemporaneous and lagged values of other variables in the VARMA system. This feature facilitates the construction of algorithms which, from the perspective of macroeconomic modeling, are efficacious in that they do not use AR approximations at any stage. Algorithms that are applicable to both asymptotically stationary and unit-root, partially nonstationary (cointegrated) time series models are presented. A sequence of lemmas and theorems show that the algorithms are based on calculations that yield strongly consistent estimates. The inferential potential of the techniques are illustrated using an example drawn from the real business cycle literature.

---

*Corresponding address: Don Poskitt, Department of Econometrics and Business Statistics, Monash University, Victoria 3800, Australia Tel.:+61-3-9905-9378; fax:+61-3-9905-5474.
   E-mail address: Don.Poskitt@Buseco.monash.edu.au

1 Introduction

Since the appearance of the seminal work of Sims (1980) on the relationship between abstract macroeconomic variables and stylized facts as represented by statistical time series models, vector autoregressive (VAR) models of the form

$$\mathbf{A}(B)\mathbf{y}_t = \mathbf{u}_t, \ t = 1, \ldots, T, \tag{1.1}$$

have become the cornerstone of much macroeconomic modeling. In equation (1.1) the vector $\mathbf{y}_t = (y_{1t}, \ldots, y_{vt})'$ denotes a v component observable process. The $v \times v$ matrix operator $\mathbf{A}(z) = \mathbf{A}_0 + \mathbf{A}_1 z^1 + \cdots + \mathbf{A}_p z^p$ in the backward shift or lag operator B, *viz.* $B\mathbf{y}_t = \mathbf{y}_{t-1}$, determines the basic evolutionary properties of the observed process $\mathbf{y}_t$ and the stochastic disturbance, $\mathbf{u}_t = (u_{1t}, \ldots, u_{vt})'$, which is unobserved, determines how chance or random influences enter the system.

Apart from their use as the main tool in numerous multivariate macroeconomic forecasting applications (as in Doan, Litterman and Sims, 1984), VARs have found broad application as the foundation of much dynamic macroeconomic modeling. They are used to study long-run equilibrium behaviour, with researchers investigating vector error correction models constructed from VARs fitted to macroeconomic time series (following Engle and Granger, 1987). In structural VAR (SVAR) models, VARs coupled with restrictions derived from economic theory are used to examine the effects of structural shocks on key macroeconomic variables (see Christiano, Eichenbaum and Vigfusson, 2006, for a recent contribution). In dynamic stochastic general equilibrium (DSGE) models, VARs are used as auxiliary models for indirect estimation of the DSGE model parameters (Smith, 1993), and to provide approximations to the solutions of DSGE models that have been expanded around their steady state (Del Negro and Schrfheide, 2004).

This ubiquitous use of VARs has occurred despite their limitations being well known. First, VAR specifications form an unattractive class of models for modeling macroeconomic variables since they are not closed under aggregation, marginalization or the presence of measurement error, see Fry and Pagan (2005) and Lütkepohl (2005). Secondly, economic models often imply that the observed processes have a vector autoregressive moving average (VARMA) representation with a non-trivial moving average component, as in Cooley and Dwyer (1998), and, more recently, Fernández-Villaverde, Rubio-Ramírez, Sargent and Watson (2007).

In order to expand the representation in (1.1) into the more general VARMA class, let us assume that $\mathbf{u}_t$ is a full rank, zero mean, p-dependent stationary process with covariance $E[\mathbf{u}_t \mathbf{u}'_{t+\tau}] = \mathbf{\Gamma}_\xi(\tau) = \mathbf{\Gamma}_\xi(-\tau)'$, $\tau = 0, \pm 1, \pm 2, \ldots, p$. This implies the existence of a sequence of zero mean, uncorrelated random variables $\boldsymbol{\varepsilon}_t$, defined on the same probability space as $\mathbf{u}_t$, such that $\mathbf{u}_t = \mathbf{M}(B)\boldsymbol{\varepsilon}_t$, $t = 1, \ldots, T$, where $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t] = \mathbf{\Sigma} > 0$ and, without loss of generality, the $v \times v$ matrix operator $\mathbf{M}(z) = \mathbf{M}_0 + \mathbf{M}_1 z^1 + \cdots + \mathbf{M}_p z^p$ satisfies $\det(\mathbf{M}(z)) \neq 0$, $|z| < 1$ (see Hannan, 1971, Theorem 10' and the associated discussion). Substituting $\mathbf{u}_t = \mathbf{M}(B)\boldsymbol{\varepsilon}_t$

into equation (1.1) gives us the VARMA form

$$\mathbf{A}(B)\mathbf{y}_t = \mathbf{M}(B)\boldsymbol{\varepsilon}_t. \tag{1.2}$$

The process $\mathbf{y}_t$ is assumed to evolve over the time period $t = 1, \ldots, T$, according to the specification given in (1.2), starting from initial values given by $\mathbf{y}_t = \boldsymbol{\varepsilon}_t = \mathbf{0}$, $t \leq 0$. The stochastic behaviour of $\mathbf{y}_t$ is now clearly dependent on the operator pair $[\mathbf{A}(z) : \mathbf{M}(z)]$, with random variation induced by the random disturbances, or shocks, $\boldsymbol{\varepsilon}_t$. More formally, it will be assumed that the disturbances, or innovations, possess the following probability structure:

Assumption 1 *The process $\boldsymbol{\varepsilon}_t = \big(\varepsilon_{1t}, \ldots, \varepsilon_{vt}\big)'$ is a stationary, ergodic, martingale differ-ence sequence. Thus if F_t denotes the σ-algebra generated by $\boldsymbol{\varepsilon}(s)$, $s \leq t$, then $E[\boldsymbol{\varepsilon}_t \mid F_{t-1}] = \mathbf{0}$. Furthermore, $E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \mid F_{t-1}] = \boldsymbol{\Sigma} > 0$ and $E[\varepsilon_{jt}^k] < \infty$, $j = 1, \ldots, v$, $k \geq 2$.*

In situations where the theoretical background gives rise to a VARMA model it might be expected that a VAR of high order could be used to approximate the true VARMA structure. Results in the recent literature suggest, however, that such an approach could be fraught with difficulties. Conditions under which the economic shocks and impulse responses from an economic model are matched by those of a (S)VAR are examined in Fernández-Villaverde et al. (2007), and Chari, Kehoe and McGrattan (2007) conclude from their analysis that the currently available data is prohibitive, leading to VARs that have too short a lag length and that provide poor approximations and unreliable inferences. For a simulated model that has both DSGE elements and data dynamics Kapetanios, Pagan and Scott (2007) suggest that a sample of 30000 observations with a VAR of order 50 is required to adequately capture the effect of some of the structural shocks. Ravenna (2007) also points out that using a VAR to characterize the dynamics of a model that in truth leads to a VARMA structure can be misleading, and warns researchers to be cautious when relying on evidence from VARs to build such models.

Given that the limitations and pitfalls of VARs for macroeconomic analysis have been well documented, one might imagine that applied macroeconomic researchers would have been compelled to consider implementing VARMA models instead. Practitioners appear to have been reluctant to embrace VARMA models however. The reason for this reluctance is, perhaps, that the complexities associated with the identification and estimation of VARMA models stand in sharp contrast to the ease and accessibility of VARs.

Multivariate time series models have, of course, been given considerable attention in the past and accounts of many of the methods and techniques available are given in Hannan and Deistler (1988) and Lütkepohl (2005), for example. Nevertheless, the question of how best to determine the internal structure of a VARMA model in a direct and straightforward manner has not been completely resolved. Two techniques of identification predominate:

1. The scalar-component methodology pioneered by Tiao and Tsay (1989), and further developed in Athanasopoulos and Vahid (2008). This method uses an adaptation of the canonical correlation analysis introduced in Akaike (1974b) to detect various linear dependencies implied by different structures. It relies on the solution of different

eigenvalue problems and solves the underlying multiple decision problem via a sequence of hypothesis tests;

2. The echelon form methodology developed in Hannan and Kavalieris (1984) and Poskitt (1992). In this approach the coefficients of an VARMA model expressed in echelon canonical form are estimated and the associated Kronecker indices determined using regression techniques and model selection criteria, *à la* AIC (Akaike, 1974a), BIC (Schwarz, 1978) or HQ (Hannan and Quin, 1979).

An illuminating exposition of the similarities and differences between scalar-component models and echelon forms is given in Tsay (1991), and Athanasopoulos, Poskitt and Vahid (2010) present a detailed analysis and comparison of these two techniques, highlighting the relative merits and advantages of each method (*c.f.* Nsiri and Roy, 1992). The lack of a single well-defined multivariate parallel to the classical Box-Jenkins ARMA methodology for univariate time series has, no doubt, discouraged researchers from employing VARMAs in practice, despite the fact that "*While VARMA models involve additional estimation and identification issues, these complications do not justify systematically ignoring these moving average components,* – – – – –." (Cooley and Dwyer, 1998). The broad aim of this paper is to fill this gap and operationalize the use of VARMA models to the point where they can be routinely employed as part of the basic toolkit of the applied macroeconomist.

The paper develops a coherent methodology for identifying and estimating VARMA models that can be fully automated. The approach adopted is to construct a modification of the echelon form methodology using a new technique to determine the Kronecker invariants. The scalar-components method is not considered here, firstly, because it is not amenable to automation in a manor similar to that used for VARs as is the echelon form methodology. Secondly, given the significance of cointegration in the practical analysis of economic and financial time series we wish to investigate unit-root nonstationary cointegrated systems and examine the consequences of applying our methods to identify cointegrated VARMA structures. Extensions of the echelon form methodology to cointegrated VARMA models have been analysised in Lütkepohl and Claessen (1997), Bartel and Lütkepohl (1998) and Poskitt (2003, 2006) (See also Lütkepohl, 2005, Chapter 14.), but to our knowledge similar extensions of the scalar-components methodology to cointegrated processes are not currently available.

In both the scalar-components and the echelon form methodologies the initial step is to fit a high-order VAR, the associated residuals are then used as plug in estimates for the unknown innovations in subsequent stages of the analysis. The applied macroeconomic literature referred to earlier questions the practical efficacy of using long VAR approximations, however, and intimates that the quality of the VAR innovations estimates is likely to be poor. Moreover, Poskitt (2005) presents theoretical arguments showing why the use of the first stage VAR residuals can lead to serious overestimation of the VARMA orders. A novel feature of the inferential procedure developed here is that it does not require the use of autoregressive approximations, thereby circumventing any problems that might be inherent in using a VAR in a macroeconomic modeling context.

The paper is organised as follows. The following section defines the the inverse echelon form and Kronecker invariants. Section 3 analyzes a single equation canonical representation that forms the basis of the identification of the Kronecker invariants. An Algorithm for the identification of the Kronecker invariants of a stationary $ARMA$ process is then presented in Section 4. Section 5 gives theoretical results stating conditions under which almost sure convergence of the estimated values to the true Kronecker invariants can be achieved. Section 6 shows how the canonical representation considered in Section 3 can be adapted to allow for cointegrated processes and Section 7 then presents a modification of the identification procedure that gives rise to a strongly consistent model selection process that is applicable to cointegrated processes. The eighth section of the paper presents the theoretical results underpinning the technique outlined in Section 7. Section 9 presents a brief conclusion and illustrates the underlying ideas and results discussed in the paper by examining the impulse responses derived from a real business cycle model.

2 The Inverse Echelon Form and Kronecker Invariants

Before continuing let us establish some additional notational conventions and assumptions. The order of $[\mathbf{A}(z) : \mathbf{M}(z)]$ is defined as $p = \max_{1 \le i \le v} n_i$ where, for $i = 1, \ldots, v$, $n_i = \delta_i [\mathbf{A}(z) : \mathbf{M}(z)]$ denotes the polynomial degree of the ith row of $[\mathbf{A}(z) : \mathbf{M}(z)]$. The integers n_r, $r = 1, \ldots, v$, are called the Kronecker indices. The Kronecker indices determine the lag structure of the, so called, inverse echelon form, which is characterized by an operator pair $[\mathbf{A}(z) : \mathbf{M}(z)]$ with polynomial elements that satisfy for all $r, c = 1, \ldots, v$,

$$
\begin{aligned}
&\text{(i)} \quad a_{rc,0} = m_{rc,0}\,, \\
&\text{(ii)} \quad m_{rr}(z) = 1 + m_{rr,1}z + \ldots + m_{rr,n_r}z^{n_r}\,, \\
&\qquad m_{rc}(z) = m_{rc,n_r-n_{rc}+1}z^{n_r-n_{rc}+1} + \ldots + m_{rc,n_r}z^{n_r} \quad \text{and} \\
&\text{(iii)} \quad a_{rc}(z) = a_{rc,0} + a_{rc,1}z + \ldots + a_{rc,n_r}z^{n_r}\,, \tag{2.1}
\end{aligned}
$$

where

$$
n_{rc} = \begin{cases} \min(n_r + 1, n_c) & r \ge c\,, \\ \min(n_r, n_c) & r < c\,. \end{cases}
$$

The restrictions implicit in (2.1) differ from those commonly found in the literature on echelon forms, see Hannan and Deistler (1988, §2.5) for detailed a discussion of the conventional case and Lütkepohl (2005, §14.2.2) for the structure in (2.1). Conditions (2.1)(i)&(ii) imply that the standard normalization $\mathbf{A}_0 = \mathbf{M}_0$, with unit leading diagonal, is imposed, but (2.1)(ii) implies that additional exclusion constraints are placed upon lower order coefficients of $\mathbf{M}(z)$ according to the relative lag lengths, rather than $\mathbf{A}(z)$. The latter feature arises because the inverse echelon form is constructed from the mapping $[\mathbf{A}(z) : \mathbf{M}(z)] \mapsto \mathbf{\Psi}(z)$ defined by $\mathbf{M}(z)\mathbf{\Psi}(z) = \mathbf{A}(z)$, wherein the coefficients $\mathbf{\Psi}_0, \mathbf{\Psi}_1, \mathbf{\Psi}_2, \ldots$ are derived from the

recursive relationships

$$\sum_{j=0}^{i} \mathbf{M}_j \mathbf{\Psi}_{i-j} = \mathbf{A}_i, \quad i = 0, \ldots, p, \text{ and}$$

$$\sum_{j=0}^{p} \mathbf{M}_j \mathbf{\Psi}_{i-j} = \mathbf{0}, \quad i = p+1, \ldots . \tag{2.2}$$

Note that $\mathbf{\Psi}_0 = \mathbf{I}$ and $\|\mathbf{\Psi}_i\| < \infty$, $i = 0, 1, \ldots$, where $\|\mathbf{\Psi}_j\|^2 = \text{tr}\mathbf{\Psi}_j\mathbf{\Psi}_j'$, the Euclidean norm. If $\det \mathbf{M}(z) \neq 0$, $|z| \leq 1$ then $\|\mathbf{\Psi}_i\| \to 0$ at an exponential rate as $i \to \infty$ and the power series $\mathbf{\Psi}(z) = \lim_{N \to \infty} \sum_0^N \mathbf{\Psi}_i z^i$ will be convergent for $|z| \leq 1$. The nomenclature is based on the fact that it is the mapping obtained via (2.2) which allows us to invert the VARMA representation and express the innovation process in terms of the model parameters and the observables, namely $\boldsymbol{\varepsilon}_t = \sum_{j=0}^{t-1} \mathbf{\Psi}_j \mathbf{y}_{t-j}$. If we let $ARMA_E(\nu)$ denote the class of all VARMA models in inverse echelon form with multi-index $\nu = \{n_1, \ldots, n_v\}$, then $ARMA_E(\nu)$ defines a canonical structure for the set of VARMA models with McMillan degree $m = \sum_{i=1}^{v} n_i$.

Assumption 2 *The pair $[\mathbf{A}(z) : \mathbf{M}(z)]$ are (left) coprime and $[\mathbf{A}(z) : \mathbf{M}(z)] \in ARMA_E(\nu)$. It will be supposed that neither $\det \mathbf{A}(z)$ or $\det \mathbf{M}(z)$ is identically zero and that the determinants of the polynomial matrices $\mathbf{A}(z)$ and $\mathbf{M}(z)$ satisfy $\det \mathbf{A}(z) \neq 0$ $|z| < 1$ and $\det \mathbf{M}(z) \neq 0$, $|z| \leq 1$.*

Note that Assumption 2 allows for the possibility that $\mathbf{A}(z)$ has zeroes on the unit circle. Let us assume that $\det \mathbf{A}(z)$ has $\zeta \leq v$ roots of unity, all other zeroes lie outside the unit circle, and that the individual series $y_{it}, i = 1, \ldots, v$, are asymptotically stationary after first differencing, i.e., $\triangle \mathbf{y}_t = (1 - B)\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$, $t = 1, \ldots, T$, is asymptotically stationary. Then the process $\mathbf{y}_t$ is non-stationary and cointegrated. We will deal with cointegrated processes in detail below, having first examined the asymptotically stationary case. For the stationary case the condition on the zeroes of $\mathbf{A}(z)$ in Assumption 2 is strengthened to $\det \mathbf{A}(z) \neq 0$, $|z| \leq 1$. We will refer to the strengthened version of Assumption 2 as Assumption 2$'$.

The Kronecker indices are not invariant with respect to an arbitrary reordering of the elements of $\mathbf{y}_t$ and to this extent the inverse echelon canonical form is only unique modulo such rotations. The variables in $\mathbf{y}_t = (y_{1t}, \ldots, y_{vt})'$ can be permuted, however, such that the Kronecker indices of $(y_{r(1)t}, \ldots, y_{r(v)t})'$ are arranged in descending order, $n_{r(1)} \geq n_{r(2)} \geq \cdots \geq n_{r(v)}$, where $r(j), j = 1, \ldots, v$, denotes a permutation of $1, \ldots, v$ that induces the ordering. The $r(j), j = 1, \ldots, v$, are unique modulo rotations that leave the ordering $n_{r(1)} \geq \cdots \geq n_{r(v)}$ unchanged and $(r(j), n_{r(j)}), j = 1, \ldots, v$, are referred to as the Kronecker invariants. When expressed in terms of the Kronecker invariants not only is the representation of the system in inverse echelon form canonical but the coefficient matrix $\mathbf{A}_0 = \mathbf{M}_0$ is lower triangular and the individual variables $y_{r(j)t}, j = 1, \ldots, v$ are uniquely characterized.

In practice, of course, the Kronecker invariants will not be known and we wish to consider identifying them in the sense of estimating or determining them from the data. Moreover, given that the numbering of the variables in $\mathbf{y}_t$ assigned by the practitioner is arbitrary,

identification of the Kronecker invariants involves the determination of not only the values of $n_{r(1)} \geq n_{r(2)} \geq \cdots \geq n_{r(v)}$, but also the permutation $(r(1), \ldots, r(v))'$ of the labels $(1, \ldots, v)'$ attached to the variables. At the risk of getting ahead of ourselves, suppose that we know that $\max\{n_1, \ldots, n_v\} \leq h$. We might contemplate examining all $ARMA$ structures in the set $\{ARMA_E(\nu) : \nu \in \{\nu = (n_1, \ldots, n_v) : 0 \leq n_r \leq h, \ r = 1, \ldots, v\}\}$. If a full search over all such structures were to be conducted then a total of $(h+1)^v$ specifications would have to be examined; if $v = 5$ and $h = 12$, say, this means estimating 371293 different $ARMA_E(\nu)$ models. This brings us face to face with the curse of dimensionality.

Considerable savings can be made, however, if the pairs $(r(j), n_{r(j)})$ associated with $y_{r(j)t}$, $j = 1, \ldots, v$, can be identified on a variable by variable, or equivalently, equation by equation, basis. Identification of the Kronecker invariants equation by equation involves examining $v(h + 1)$ different specifications at most; if $v = 5$ and $h = 12$ this gives an upper bound of 65, rather than the previous total of 371293. To determine the Kronecker invariants variable by variable, however, we need to construct from the $ARMA_E(\nu)$ system representation a univariate specification for $y_{r(j)t}$ that allows the Kronecker invariant pairs $(r(j), n_{r(j)})$, $j = 1, \ldots, v$, to be appropriately isolated. We derive such a specification in the next section.

3 A Single Equation Canonical Structure

Various aspects of the relationship between VARMA models and the structure of the individual univariate series have been discussed in the literature, but none consider specifications that are suitable for our current purposes since they all convolve the individual operators in such a way as to disguise their underlying polynomial degrees. The final form (Wallis, 1977), for example, is obtained by pre-multiplying (1.2) by the adjoint of $\mathbf{A}(z)$, denoted $\mathrm{adj}\mathbf{A}(z)$, to give

$$\det \mathbf{A}(B)\mathbf{y}_t = \mathrm{adj}\mathbf{A}(B)\mathbf{M}(B)\boldsymbol{\varepsilon}_t . \tag{3.1}$$

In general, the operators on the left and right hand sides of this expression all have degree equal to m. Consequently, although (3.1) can be used to determine the McMillan degree of the overall system (see Remark 3 below), it does not yield univariate specifications suitable for the identification of the Kronecker invariants.

In order to identify $n_{r(1)} \geq n_{r(2)} \geq \cdots \geq n_{r(v)}$, we now introduce a single equation canonical structure, derived from (1.2), that does not obscure the Kronecker invariants. The single equation form depends upon the following lemma.

Lemma 3.1 *Suppose that $\mathbf{y}_t$ is an ARMA process as in* (1.2) *satisfying Assumptions 1 and 2. Then for each choice of the process $\nu_t = u_{jt}$, $j = 1, \ldots, v$, where*

$$\mathbf{u}_t = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{vt} \end{bmatrix} = \mathbf{A}(B)\mathbf{y}_t ,$$

there exists a zero mean, scalar white noise process η_t, with variance σ_η^2, defined on the same

probability space as $\mathbf{y}_t$, *such that*

$$\nu_t = \eta_t + \mu_1 \eta_{t-1} + \ldots + \mu_n \eta_{t-n},$$

where $n = n_j$ *and the coefficients* $\mu_1, \ldots, \mu_n$ *of* $\mu(z) - 1 = \sum_{s=1}^{n} \mu_s z^s$ *are such that the auto-covariance generating function of* ν_t *equals* $\sigma_\eta^2 \mu(z) \mu(z^{-1})$ *and* $\mu(z) \neq 0$, $|z| \leq 1$.

A consequence of Lemma 3.1 is that each variable in $\mathbf{y}_t$ admits a scalar ARMAX representation in which the exogenous variables are chosen from contemporaneous and lagged values of other variables in the VARMA system.

Proposition 3.1 *Let* $\mathbf{y}_t$ *be an ARMA process satisfying Assumptions 1 and 2, and suppose that the variables have been ordered (renumbered) according to the Kronecker invariants, so that, with a slight abuse of notation,* $\mathbf{y}_t = (y_{1t}, \ldots, y_{vt})' = (y_{r(1)t}, \ldots, y_{r(v)t})'$ *and* $n_j = n_{r(j)}$, $j = 1, \ldots, v$. *Then for each* $j = 1, \ldots, v$ *the jth equation in (1.2) is equivalent to a scalar ARMAX specification for* $z_t = y_{jt}$ *of the form*

$$z_t + \sum_{s=1}^{n} \alpha_s z_{t-s} + \sum_{i=1}^{j-1} \beta_i y_{it} + \sum_{\substack{i=1 \\ i \neq j}}^{v} \sum_{s=1}^{n} \beta_{i,s} y_{it-s} = \eta_t + \sum_{s=1}^{n} \mu_s \eta_{t-s}, \qquad (3.2)$$

where the order $n = n_j$. *Moreover,* $\alpha(z) = 1 + \sum_{s=1}^{n} \alpha_s z^s$, $\boldsymbol{\beta}(z) = (\beta_1 + \sum_{s=1}^{n} \beta_{1,s} z^s, \ldots, \beta_{j-1} + \sum_{s=1}^{n} \beta_{j-1,s} z^s, 0, \sum_{s=1}^{n} \beta_{j+1,s} z^s, \ldots, \sum_{s=1}^{n} \beta_{v,s} z^s)$ *and* $\mu(z) = 1 + \sum_{s=1}^{n} \mu_s z^s$ *are coprime, and the regressors* y_{it}, $i = 1, \ldots, j-1$, *and* y_{it-s}, $i = 1, \ldots, v$, $i \neq j$, $s = 1, \ldots, n$ *are predetermined relative to* z_t *in the representation (3.2).*

Two aspects of Proposition 3.1 that are of particular interest here are; (i) that the degree of the lag operators in (3.2) depends only on the value of the Kronecker invariant associated with the variable at hand, and (ii) that the contemporaneous component depends only on those variables associated with a smaller Kronecker invariant. This means that knowledge of the Kronecker index associated with $y_{r(j)t}$ tells us the lag length of all the variables appearing in the ARMAX realization of $y_{r(j)t}$, and knowing the ranking of the Kronecker index relative to the other indices, i.e. knowledge that $n_{r(j)} \geq n_{r(i)}$, $i = j+1, \ldots, v$, tells us about the recursive structure. Note that explicit knowledge of the values of the other Kronecker invariants is not required, the ARMAX specification of the $r(j)$th equation being a known function of $d_j(n) = (j-1) + (v+1)n$ parameters. These features suggest that a sensible approach to adopt to the identification of the Kronecker invariants is to search through a collection of ARMAX models for each variable supposing that the fitted order coincides with the *smallest* unknown Kronecker invariant. The details of such a procedure are presented in the following section.

4   Identification Algorithm: Stationary Case

Let $\mathbf{y}_t$, $t = 1, \ldots, T$ denote a realisation of T observations where $\mathbf{y}_t$ is an ARMA process as in (1.2) satisfying Assumptions 1 and 2'. We have already observed that identification of the Kronecker invariants involves the determination of both the values of $n_{r(1)} \geq \cdots \geq n_{r(v)}$ and

the permutation of the variables in $\mathbf{y}_t$, $\mathbf{P}(1,\ldots,v)' = (r(1),\ldots,r(v))'$ say, that results in an $ARMA_E$ form with multi-index $(n_{r(1)},\ldots,n_{r(v)})$ for $\mathbf{P}\mathbf{y}_t = (y_{r(1)t},\ldots,y_{r(v)t})'$. The following algorithm identifies the Kronecker invariants equation by equation whilst constructing $\mathbf{P}$ via a sequence of elementary row operations. The algorithm exploits the implications of Proposition 3.1 and represents an adaptation to VARMA models of an approach to the identification of the order of scalar processes first outlined in Poskitt and Chung (1996).

ALGORITHM–$ARMA_E(\nu)$.

Initialization: Set $n = 1$, $j = v$, $\mathbf{P} = \mathbf{I}$ and $\mathcal{N} = \{1,\ldots,v\}$. Compute the mean corrected values $\bar{\mathbf{y}}_t = \mathbf{y}_t - \bar{\mathbf{y}}$, $t = 1,\ldots,T$, where $\bar{\mathbf{y}} = T^{-1}\sum_{i=1}^T \mathbf{y}_t$. For each $i \in \mathcal{N}$, set $\widetilde{\sigma}_{\eta,i}^2(0)$ equal to the residual mean square from the regression of $\bar{y}_{it}$ on $\bar{y}_{kt}$, $k = 1,\ldots,v$, $k \neq i$.

while: $j \geq 1$

> **for** $i(k) \in \mathcal{N}$, $k = 1,\ldots,v$,
>
> > 1. Set $\widetilde{\mathbf{y}}_t = \mathbf{E}_{i(k),j}[\mathbf{P}\mathbf{y}_t]$, where $\mathbf{E}_{r_1,r_2}$ denotes the $v \times v$ elementary matrix that induces an interchange of rows r_1 and r_2 in $\mathbf{H}$ when postmultiplied by $\mathbf{H}$, and evaluate initial estimates of the jth scalar ARMAX form for $z_t = \bar{y}_{i(k)t}$:
> >
> > (a) Construct estimates of the nonzero coefficients in $\widetilde{\mathbf{a}}(z) = \mathbf{e}_j'\widetilde{\mathbf{A}}(z)$, the jth row of $\widetilde{\mathbf{A}}(z) = \mathbf{E}_{i(k),j}\mathbf{P}\mathbf{A}(z)\mathbf{P}'\mathbf{E}_{i(k),j}'$, by solving the equations
> >
> > $$\sum_{s=0}^n \widehat{\widetilde{\mathbf{a}}}_s \widetilde{\mathbf{C}}_y(r+s) = \mathbf{0}, \ r = n+1,\ldots,2n, \tag{4.1}$$
> >
> > for $\widehat{\widetilde{\mathbf{a}}}_0 = (\widehat{\widetilde{a}}_{1,0},\ldots,\widehat{\widetilde{a}}_{(j-1),0},1,0,\ldots,0)$, and $\widehat{\widetilde{\mathbf{a}}}_s = (\widehat{\widetilde{a}}_{1,s},\ldots,\widehat{\widetilde{a}}_{v,s})$, $s = 1,\ldots,n$, where
> >
> > $$\widetilde{\mathbf{C}}_y(r) = \widetilde{\mathbf{C}}_y(-r)' = T^{-1}\sum_{t=1}^{T-|r|} \widetilde{\mathbf{y}}_t \widetilde{\mathbf{y}}_{t-r}'$$
> >
> > for $r = 1,\cdots,T-1$. Now set $\widehat{\widetilde{\alpha}}_s = \widehat{\widetilde{a}}_{j,s}$, $s = 1,\ldots,n$, $\widehat{\widetilde{\beta}}_i = \widehat{\widetilde{a}}_{i,0}$, $i = 1,\ldots,j-1$, and $\widehat{\widetilde{\beta}}_{i,s} = \widehat{\widetilde{a}}_{i,s}$, $i = 1,\ldots,v$, $i \neq j$, $s = 1,\ldots,n$.
> >
> > (b) For $r = 1,\cdots,n$ form
> >
> > $$\widetilde{C}_v(r) = \widetilde{C}_v(-r) = \sum_{s=0}^n\sum_{u=0}^n \widehat{\widetilde{\mathbf{a}}}_s\widetilde{\mathbf{C}}_y(r+s-u)\widehat{\widetilde{\mathbf{a}}}_u$$
> >
> > $$= \int_{-\pi}^{\pi} \widehat{\widetilde{\mathbf{a}}}(\omega)\widetilde{\mathbf{I}}_y(\omega)\widehat{\widetilde{\mathbf{a}}}(\omega)^* \exp(-\imath\omega r)d\omega \tag{4.2}$$
> >
> > where $\widetilde{\mathbf{a}}(\omega) = \widetilde{\mathbf{a}}_0 + \widetilde{\mathbf{a}}_1\exp(\imath\omega) + \cdots + \widetilde{\mathbf{a}}_n\exp(\imath\omega n) = \widetilde{\mathbf{a}}(z)\big|_{z=e^{\imath\omega}}$ and
> >
> > $$\widetilde{\mathbf{I}}_y(\omega) = \frac{1}{2\pi}\sum_{s=-T+1}^{T-1} \widetilde{\mathbf{C}}_y(s)\exp(\imath\omega).$$
> >
> > Set
> >
> > $$\widehat{\widetilde{S}}_v(\omega) = \frac{1}{2\pi}\sum_{s=-n}^n \widetilde{C}_v(s)\exp(\imath\omega s). \tag{4.3}$$

Compute estimates $\widehat{\widetilde{\mu}}_s$, $s = 1, \ldots, n$, of the coefficients in the scalar moving average representation of $v_t = \widetilde{u}_{jt}$, where $\widetilde{u}_{jt} = \mathbf{e}_j' \widetilde{\mathbf{M}}(B)\varepsilon_t$ and $\widetilde{\mathbf{M}}(z) = \mathbf{E}_{i(k),j} \mathbf{P} \mathbf{M}(z)$, by solving the equation system

$$\sum_{s=0}^{n} \widehat{\widetilde{\mu}}_s \widetilde{C}i_v(l + s) = 0, \; l = 1, \ldots, n$$

where

$$\widetilde{C}i_v(r) = \int_{-\pi}^{\pi} \frac{\widehat{\widetilde{\mathbf{a}}}(\omega) \widetilde{\mathbf{I}}_y(\omega) \widehat{\widetilde{\mathbf{a}}}(\omega)^*}{\widehat{\widetilde{S}}_v(\omega)^2} \exp(-\imath \omega r) d\omega \,. \tag{4.4}$$

2. Compute a pseudo maximum likelihood estimate (pseudo MLE) of the innovation variance $\widetilde{\sigma}_\eta^2$.

(a) Form the $v + 1$ vector sequence $(\widehat{\widetilde{\boldsymbol{\xi}}}_t', \widehat{\widetilde{\varphi}}_t,)'$ by solving

$$\begin{bmatrix} \widehat{\widetilde{\boldsymbol{\xi}}}_t \\ \widehat{\widetilde{\varphi}}_t \end{bmatrix} + \sum_{s=1}^{n} \widehat{\widetilde{\mu}}_s \begin{bmatrix} \widehat{\widetilde{\boldsymbol{\xi}}}_{t-s} \\ \widehat{\widetilde{\varphi}}_{t-s} \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{y}}_t \\ \widehat{\widetilde{\eta}}_t \end{bmatrix}$$

for $t = 1, \ldots, T$ where

$$\begin{aligned} \widehat{\widetilde{\eta}}_t &= \sum_{s=0}^{n} \widehat{\widetilde{\mathbf{a}}}_{j,s} \widetilde{\mathbf{y}}(t - s) - \sum_{s=1}^{n} \widehat{\widetilde{\mu}}_s \widehat{\widetilde{\eta}}_{t-s} \\ &= \sum_{s=0}^{n} \widehat{\widetilde{\alpha}}_s z_{t-s} + \sum_{i=1}^{j-1} \widehat{\widetilde{\beta}}_i \widetilde{y}_{it} + \sum_{\substack{i=1 \\ i \neq j}}^{v} \sum_{s=1}^{n} \widehat{\widetilde{\beta}}_{i,s} \widetilde{y}_{it-s} - \sum_{s=1}^{n} \widehat{\widetilde{\mu}}_s \widehat{\widetilde{\eta}}_{t-s} \end{aligned}$$

and the recursions are initiated at $\widehat{\widetilde{\eta}}_t = 0$ and $(\widehat{\widetilde{\boldsymbol{\xi}}}_t', \widehat{\widetilde{\varphi}}_t,)' = \mathbf{0}'$, $t \leq 0$.

(b) Set $\widetilde{c}_{\widehat{\eta}}(0) = T^{-1} \sum_t^T \widehat{\widetilde{\eta}}_t^2$ and calculate

$$\widetilde{\mathbf{c}}_{\widehat{\eta}\widehat{\xi}}(s) = T^{-1} \sum_{t=s+1}^{T} \widehat{\widetilde{\eta}}_t \widehat{\widetilde{\boldsymbol{\xi}}}_{t-s}' \quad \text{and} \quad \widetilde{c}_{\widehat{\eta}\widehat{\varphi}}(s) = T^{-1} \sum_{t=s+1}^{T} \widehat{\widetilde{\eta}}_t \widehat{\widetilde{\varphi}}_{t-s} \,.$$

Now compute the mean squared error

$$\widehat{\widetilde{\sigma}}_\eta^2(n) = \widetilde{c}_{\widehat{\eta}}(0) + \sum_{s=0}^{n} \Delta \widehat{\widetilde{\mathbf{a}}}_s \widetilde{\mathbf{c}}_{\widehat{\eta}\widehat{\xi}}(s)' - \sum_{s=1}^{n} \Delta \widehat{\widetilde{\mu}}_s \widetilde{c}_{\widehat{\eta}\widehat{\varphi}}(s)$$

where $\Delta \widehat{\widetilde{\mathbf{a}}}_s$, $s = 0, \ldots, n$, and $\Delta \widehat{\widetilde{\mu}}_s$, $s = 1, \ldots, n$, denote the coefficient values obtained from the Toeplitz regression of $\widehat{\widetilde{\eta}}_t$ on $-\widehat{\widetilde{\xi}}_{it}$, $i = 1, \ldots, j-1$, and $-\widehat{\widetilde{\boldsymbol{\xi}}}_{t-s}$, $s = 1, \ldots, n$, and $\widehat{\widetilde{\varphi}}_{t-s}$, $s = 1, \ldots, n$.

3. Apply model selection rule:

(a) Evaluate the criterion function

$$IC_T(n) = T \log \widehat{\widetilde{\sigma}}_\eta^2(n) + p_T(d_j(n))$$

where the penalty term $p_T(d_j(n)) > 0$ is a real valued function, monotonically increasing in $d_j(n)$ and non-decreasing in T.

(b) **if** $IC_T(n) > IC_T(n-1)$;
    set $r(j) = i(k)$; $n_{r(j)} = n - 1$; and
    update $\mathbf{P} = \mathbf{E}_{i(k),j}\mathbf{P}$; $\mathcal{N} = \mathcal{N} \setminus i(k)$; $j = j - 1$.
    end

end for $i(k) \ni \mathcal{N}$.

if $n < h_T$;
    increment $n = n + 1$.
    else
    for $i(k) \in \mathcal{N}$, $k = 1, \ldots, v$,
        set $n_{r(j)} = n$; $r(j) = i(k)$; and
        update $\mathbf{P} = \mathbf{E}_{i(k),j}\mathbf{P}$; $\mathcal{N} = \mathcal{N} \setminus i(k)$; $j = j - 1$.
        end for $i(k) \ni \mathcal{N}$.
    end

end when $j = 0$.

Some remarks on the algorithm's rationale and numerical implementation are in order:

REMARK 1. The first step of the algorithm is designed to provide first stage consistent estimates of the parameters in the scalar ARMAX representation in (3.2). Step 1(a) is based on the fact that from Proposition 3.1 it follows that $\widetilde{\mathbf{y}}_{t-n-s}$, $s = 1, \ldots, n$, are orthogonal to $\mathbf{e}'_j\widetilde{\mathbf{M}}(B)\boldsymbol{\varepsilon}_t = \eta_t + \mu_1\eta_{t-1} + \ldots + \mu_n\eta_{t-n}$, the scalar moving average corresponding to the right hand side of the $r(j)$th equation. Thus, post multiplying by $\widetilde{\mathbf{y}}'_{t-n-s}$, taking expectations, and writing $\widetilde{\boldsymbol{\Gamma}}_y(r) = E[\widetilde{\mathbf{y}}_t\widetilde{\mathbf{y}}'_{t-r}]$ for the theoretical autocovariance function of $\widetilde{\mathbf{y}}_t$, remembering that in Proposition 3.1 the variables are assumed to be ordered according to the Kronecker invariants, we find that

$$\sum_{s=0}^{n} \widetilde{\mathbf{a}}_s\widetilde{\boldsymbol{\Gamma}}_y(r + s) = \mathbf{0}, \ r = n + 1, \ldots, 2n. \tag{4.5}$$

We can see that expression (4.1) forms an empirical counterpart to (4.5) in which the variables have been appropriately permuted. Given that $\widetilde{\mathbf{C}}_y(r)$ is a strongly consistent estimator of $\widetilde{\boldsymbol{\Gamma}}_y(r)$ it follows immediately that if the Kronecker invariant pairs $(r(i), n_{r(i)})$, $i = j, \ldots, v$, are correctly specified, then $\widehat{\widetilde{\alpha}}(z)$ and $\widehat{\widetilde{\boldsymbol{\beta}}}(z)$ will yield strongly consistent estimates of $\widetilde{\alpha}(z)$ and $\widetilde{\boldsymbol{\beta}}(z)$ respectively.

A corollary of the consistency of $\widehat{\widetilde{\alpha}}(z)$ and $\widehat{\widetilde{\boldsymbol{\beta}}}(z)$ is that

$$\int_{-\pi}^{\pi} \widehat{\widetilde{\mathbf{a}}}(\omega)\widetilde{\mathbf{I}}_y(\omega)\widehat{\widetilde{\mathbf{a}}}(\omega)^* \exp(-\imath\omega r)d\omega = \int_{-\pi}^{\pi} \widetilde{\mathbf{a}}(\omega)\widetilde{\mathbf{S}}_y(\omega)\widetilde{\mathbf{a}}(\omega)^* \exp(-\imath\omega r)d\omega + o(1)$$

almost surely (a.s.) wherein $\widetilde{\mathbf{S}}_y(\omega)$ denotes the spectral density of $\widetilde{\mathbf{y}}_t$. Given that $\widetilde{\mathbf{S}}_y(\omega) = (2\pi)^{-1}\widetilde{\mathbf{K}}(\omega)\boldsymbol{\Sigma}\widetilde{\mathbf{K}}(\omega)^*$ where $\widetilde{\mathbf{K}}(\omega) = \widetilde{\mathbf{A}}(\omega)^{-1}\widetilde{\mathbf{M}}(\omega)$ it follows that the quadratic form

$$\begin{aligned}
\widetilde{\mathbf{a}}(\omega)\widetilde{\mathbf{S}}_y(\omega)\widetilde{\mathbf{a}}(\omega)^* &= 2\pi)^{-1}\mathbf{e}'_j\widetilde{\mathbf{M}}(\omega)\boldsymbol{\Sigma}\widetilde{\mathbf{M}}(\omega)^*\mathbf{e}_j \\
&= (2\pi)^{-1}\sigma_\eta^2|\widetilde{\mu}(\omega)|^2
\end{aligned}$$

and hence that the autocovariance estimate computed in Step 1(b) at (4.2) is consistent for the autocovariance of $\nu_t = \widetilde{u}_{jt}$. The spectral estimator $\widehat{\widetilde{S}}_v(\omega)$ computed at (4.3) is therefore consistent for $\widetilde{S}_v(\omega)$, the spectrum of ν_t. Similarly,

$$\int_{-\pi}^{\pi} \frac{\widehat{\widetilde{\mathbf{a}}}(\omega)\widetilde{\mathbf{I}}_y(\omega)\widehat{\widetilde{\mathbf{a}}}(\omega)^*}{\widehat{\widetilde{S}}_v(\omega)^2} \exp(-\imath\omega r)d\omega = \int_{-\pi}^{\pi} \frac{1}{\widetilde{S}_v(\omega)} \exp(-\imath\omega r)d\omega + o(1) \quad \text{a.s.}$$

so the $\widetilde{C}i_v(r)$ in (4.4) yield consistent estimates of the corresponding inverse autocovariances, the coefficients in the Fourier expansion of $\widetilde{S}_v(\omega)^{-1} = 2\pi/\sigma_\eta^2|\widetilde{\mu}(\omega)|^2$. From the latter it follows that $\widehat{\widetilde{\mu}}_s$, $s = 1, \ldots, n$, provide consistent estimates of the coefficients in $\widetilde{\mu}(z)$. Detailed particulars of the arguments underlying this heuristic rationale are presented below.

REMARK 2. It is a straightforward to show that $\widehat{\widetilde{\alpha}}(z)$ is equivalent to the estimator obtained using $\widetilde{\mathbf{y}}_{t-n-s}$, $s = 1, \ldots, n$, as instruments in an exactly–identified instrumental variables regression estimate of the autoregressive coefficients of the $r(j)$th equation. Given that $\widetilde{\mathbf{y}}_{t-n-s}$ for all $s \geq 1$ are admissible instruments, we could increase the number of instruments by taking $\widetilde{\mathbf{y}}_{t-n-s}$, $s = 1, \ldots, h$, where $h > n$, and calculate the over–identified instrumental variables estimator, $\widehat{\widetilde{\alpha}}_h(z)$ say. Stoica, Soderstrom and Friedlander (1985) examine this possibility and show that the asymptotic variance of $\widehat{\widetilde{\alpha}}_h(z)$ is monotonically decreasing in h and converges to a finite limit, and hence that asymptotic efficiency is achieved by letting $h \to \infty$ as $T \to \infty$. It is envisaged that in practice $\widehat{\widetilde{\alpha}}_h(z)$ will be evaluated with $h = a \log T$, $a > 0$, but for ease of exposition and notational simplicity we will confine our attention to $\widehat{\widetilde{\alpha}}(z)$ here.

REMARK 3. Under Gaussian assumptions $-\frac{T}{2}\log\frac{1}{T}\sum_t^T \widetilde{\eta}_t^2$, where $\widetilde{\mu}(B)\widetilde{\eta}_t = \widetilde{\alpha}(B)z_t + \widetilde{\beta}(B)'\widetilde{\mathbf{y}}_t = \widetilde{\mathbf{a}}(B)\widetilde{\mathbf{y}}_t$, forms an approximation to the kernel of the marginal log likelihood of the scalar ARMAX specification, concentrated with respect to σ_η^2. Recalling the convention about values before $t = 1$, we find that

$$\frac{\partial\widetilde{\eta}_t}{\partial\alpha_s} = \widetilde{\xi}_{j(t-s)},$$
$$\frac{\partial\widetilde{\eta}_t}{\partial\beta_{is}} = \widetilde{\xi}_{i(t-s)} \quad \text{and}$$
$$\frac{\partial\widetilde{\eta}_t}{\partial\mu_s} = -\widetilde{\varphi}_{t-s},$$

and hence that

$$\frac{\partial\sum_t^T\widetilde{\eta}_t^2}{\partial\mathbf{a}_j'} = 2\sum_{t=1}^T \widetilde{\boldsymbol{\xi}}_{t-j}\widetilde{\eta}_t \quad \text{and} \quad \frac{\partial\sum_t^T\widetilde{\eta}_t^2}{\partial\mu_j} = -2\sum_{t=1}^T \widetilde{\varphi}_{t-j}\widetilde{\eta}_t.$$

Thus Step 2 may be viewed as a Gauss-Newton iteration designed to minimize $T^{-1}\sum_t^T \widetilde{\eta}_t^2$, in line with the univariate procedure of Hannan and Rissanen (1982), only now the calculations are initiated with the consistent parameter estimates provided by $\widehat{\widetilde{\mathbf{a}}}(z)$ and $\widehat{\widetilde{\mu}}(z)$. Revised parameter estimates are constructed as $\widehat{\widetilde{\alpha}}_s + \Delta\widehat{\widetilde{\alpha}}_s$, $\widehat{\widetilde{\beta}}_{i,s} + \Delta\widehat{\widetilde{\beta}}_{i,s}$ and $\widehat{\widetilde{\mu}}_s + \Delta\widehat{\widetilde{\mu}}_s$, $s = 1, \cdots, n$, where the parameter adjustments are given by the regression coefficients in the regression of $\widehat{\widetilde{\eta}}_t$ on $-\widehat{\widetilde{\xi}}_{it}$, $i = 1, \ldots, j-1$, and $-\widehat{\widetilde{\boldsymbol{\xi}}}_{t-s}$ and $\widehat{\widetilde{\varphi}}_{t-s}$, $s = 1, \ldots, n$. The iteration can then be

repeated until convergence occurs, if so desired. Because $T^{-1} \sum_t^T \widetilde{\eta}_t^2$ is 'quasi–quadratic' it seems likely that for large values of T no more than two or three iterations will be required for convergence. For theoretical purposes we will therefore assume that the minimum residual mean square has been achieved, although in order to maintain a closed form expression for the algorithm we have chosen to express the pseudo MLE via a single iteration.

REMARK 4. The decision rule embodied in Step 3 leads to the identification of the Kronecker invariants as the first local minimum of $IC_T(n)$ in the interval $0 \leq n \leq h_T$, so that $IC_T(n) \geq IC_T(n_{r(j)})$, for $n < n_{r(j)}$, and either $IC_T(n_{r(j)} + 1) > IC_T(n_{r(j)})$ or $n_{r(j)} = h_T$, $j = 1, \ldots, v$. It follows that once the practitioner has designated a specification for the criterion function $IC_T(n)$, and prescribed the upper bound h_T, the identification of the Kronecker invariants is a fully automatic procedure.

Many well known information theoretic criteria, such as AIC and BIC, are encompassed by $IC_T(n)$, and in light of the extensive literature on such criteria we can anticipate that if the penalty term $p_T(d_j(n))$ is assigned appropriately then asymptotically the criterion function $IC_T(n)$ will possess a global minimum when n equals the true Kronecker index, $n_{r(j)0}$. That this is indeed the case is verified below, where it is shown that the Kronecker invariants identified by the algorithm will be strongly consistent if $p_T(d_j(n)) \to 0$ as $T \to \infty$ such that $Tp_T(d_j(n)) \to \infty$ and $\log\log T/(T \cdot p_T(d_j(n))) \to 0$. These requirements allow for a wide range of possibilities beyond the conventional AIC or BIC type penalties, suggesting that the most appropriate choice of $IC_T(n)$ may be an empirical issue.

Obviously the design parameter h_T must be assigned such that $h_T \geq \max\{n_1, \ldots, n_v\}$. This can be done by noting from the final form in (3.1) that each variable in $\mathbf{y}_t$ has a scalar $ARMA(q, q)$ representation with $q \leq m$ and $q = m$ for at least one y_{jt}, $j = 1, \ldots, v$. By applying the univariate ARMA algorithm of Poskitt and Chung (1996) to each y_{jt}, $j = 1, \ldots, v$, we can generate v estimates, $\widehat{q}_1, \ldots, \widehat{q}_v$ say. Suppose that this is done using $AIC(q)$ for the criterion function, for q over the range $0 \leq q \leq \log T^a$, $a > 1$. Setting $h_T = \max\{\widehat{q}_1, \ldots, \widehat{q}_v\}$ yields an 'estimate' of the McMillan degree and by Theorem 4.1 of Poskitt and Chung (1996) $\lim_{T \to \infty} h_T \geq m$ with probability one. Thus, h_T provides a value for the upper bound that will exceed the largest Kronecker invariant almost surely.

REMARK 5. Thus far we have expressed the computational steps of the algorithm in terms of the required statistical calculations but we have not commented on numerical implementation. Various measures can be taken to optimize the efficiency of the computations. For example, advantage can be taken of the fast Fourier transform (FFT) when evaluating the covariances and convolutions required to implement the algorithm. Thus, the covariances $\mathbf{C}_y(r)$, $r = 0, \pm 1, \ldots, \pm(T-1)$, of the raw data $\mathbf{y}_t$ can be calculated once and for all from

$$\mathbf{C}_y(r) = \frac{2\pi}{N} \sum_{s=1}^N \mathbf{I}_y(\omega_s) \exp(-\iota\omega_s r)$$

where $\omega_s = 2\pi s/N$, $s = 1, \ldots, N \geq 2T$ and $\mathbf{I}_y(\omega) = (2\pi T)^{-1}\mathbf{Z}_y(\omega)\mathbf{Z}_y(\omega)^*$ with $\mathbf{Z}_y(\omega) = \sum_{t=1}^T \mathbf{y}_t \exp(\iota\omega t)$. It is well known (Bingham, 1974) that this method takes an order of $T \log T$ operations rather than the order T^2 operations used with standard methods. The

autocovariances and periodogram ordinates of $\widetilde{\mathbf{y}}_t$ can then be determined using elementary row and column transformations, as in

$$\widetilde{\mathbf{C}}_y(r) = \mathbf{E}_{i(k),j}\mathbf{P}\mathbf{C}_y(r)\mathbf{P}'\mathbf{E}'_{i(k),j} \qquad \text{and} \qquad \widetilde{\mathbf{I}}_y(\omega) = \mathbf{E}_{i(k),j}\mathbf{P}\mathbf{I}_y(\omega)\mathbf{P}'\mathbf{E}'_{i(k),j}\,.$$

Similarly, the frequency domain expression for $\widetilde{C}_v(r)$ in (4.2) is not suitable for computation, but the integral may be replaced by an appropriate Riemann sum and evaluated via the FFT using

$$
\begin{aligned}
\widetilde{C}_v(r) &= \sum_{s=0}^{n}\sum_{u=0}^{n}\widehat{\widetilde{\mathbf{a}}}_s\widetilde{\mathbf{C}}_y(r+s-u)\widehat{\widetilde{\mathbf{a}}}_u \\
&= \frac{2\pi}{N}\sum_{s=1}^{N}\widehat{\widetilde{\mathbf{a}}}(\omega_s)\widetilde{\mathbf{I}}_y(\omega_s)\widehat{\widetilde{\mathbf{a}}}(\omega_s)^*\exp(-\imath\omega_s r)\,. \quad (4.6)
\end{aligned}
$$

Since both $\widetilde{\mathbf{I}}_y(\omega)$ and $\widehat{\widetilde{\mathbf{a}}}(\omega)$ are polynomial (*time-limited*) the use of (4.6) does not induce aliasing, whereas, replacing the integral in (4.4) by

$$\frac{2\pi}{N}\sum_{s=1}^{N}\frac{\widehat{\widetilde{\mathbf{a}}}(\omega_s)\widetilde{\mathbf{I}}_y(\omega_s)\widehat{\widetilde{\mathbf{a}}}(\omega_s)^*}{\widehat{\widetilde{S}}_v(\omega_s)^2}\exp(-\imath\omega_s r) = \sum_{j=-\infty}^{\infty}\widetilde{C}i_v(r+jN)$$

clearly results in some aliasing relative to the basic definition of $\widetilde{C}i_v(r)$. However, $\widetilde{S}_v(\omega)$ corresponds to the power spectrum of an invertible moving–average, implying that for T sufficiently large $|\widetilde{C}i_v(r)| < \kappa\lambda^{|r|}$ with probability one, where $0 < \lambda < 1$ and κ denotes a fixed constant. Thus $|\widetilde{C}i_v(u) - \sum_{j=-\infty}^{\infty}\widetilde{C}i_v(r+jN)| < 2\kappa\exp(N\log\lambda)/(1-\lambda^N)$ and the effects of aliasing will disappear asymptotically.

REMARK 6. The calculation of $\widehat{\widetilde{\alpha}}(z)$, $\widehat{\widetilde{\boldsymbol{\beta}}}(z)$ and $\widehat{\widetilde{\mu}}(z)$ are Toeplitz in nature, meaning that the matrices in the linear equations being solved have constant elements down any diagonal. This feature is particularly important in the context of $\widehat{\widetilde{\mu}}(z)$ because at Step 2 it is necessary for $\widehat{\widetilde{\mu}}(z)$ to be invertible, for otherwise the recursions forming $(\widehat{\widetilde{\boldsymbol{\xi}}}'_t, \widehat{\widetilde{\varphi}}_t,)'$ will explode. The requirement that $\widehat{\widetilde{\mu}}(z) \neq 0$, $|z| \leq 1$, is met since solving (4.4) for $\widehat{\widetilde{\mu}}(z)$ is equivalent to solving Yule-Walker equations in the inverse autocovariances. When computing $\widehat{\widetilde{\mu}}(z)$ advantage can therefore be taken of the Levinson-Durbin recursions. Indeed, we can also embed the evaluation of $\widehat{\widetilde{\alpha}}(z)$ and $\widehat{\widetilde{\boldsymbol{\beta}}}(z)$, as well as the calculations of Step 2, into appropriate multivariate Levinson-Durbin (Whittle) recursions. Details of the latter, which follow the development in Hannan and Deistler (1988, pp. 249-251), are omitted.

It is well known that the use of Toeplitz calculations can have undesirable end-effects. These effects can be ameliorated by the use of Burg-type procedures (Paulsen and Tjøstheim, 1985; Tjøstheim and Paulsen, 1983), but the use of a data-taper in conjunction with Whittle type estimators, such as those implicitly being employed here, can be equally beneficial (Dahlhaus, 1988). Moreover, the benefits obtained via a data-taper can be achieved without incurring the additional computational burden entailed in using Burg-type procedures. Given that we envisage conducting the computations using the FFT the employment of data-tapering seems natural.

5 Some Theoretical Properties

This section of the paper presents the main theorems for the asymptotically stationary case. The results provide conditions on the penalty term $p_T(d_j(n))$ assigned to the criterion function $IC_T(n)$ that will ensure that the indices obtained by implementing the above algorithm will yield consistent estimates of the Kronecker invariants. Recall that identification of the Kronecker invariants also involves the determination of the permutation $(r(1), \ldots, r(v))'$ of the original labels $(1, \ldots, v)'$ attached to the variables. In what follows we will let $r(q)_T$, $q = 1, \ldots, v$, denote the reordering of $r = 1, \ldots, v$ induced by $n_{r(j)T}$, $j = 1, \ldots, v$, and we will employ the labels $r(1)^0, \ldots, r(v)^0$ for the reordering associated with true Kronecker invariants $n^0_{r(j)}$, $j = 1, \ldots, v$.

Theorem 5.1 *Suppose that $\mathbf{y}_t$ is an ARMA process satisfying Assumptions 1 and 2′, and let $\{r(j)_T, n_{r(j)T}\}$, $j = 1, \ldots, v$, denote the Kronecker invariant pairs obtained obtained when employing the above algorithm with $p_T(d_j(n))$ a possibly stochastic function of n and T. Then:*

(i) If $(r(i)_T, n_{r(i)T}) = (r(i)^0, n^0_{r(i)})$, $i = q + 1, \ldots, v$, and $p_T(d_q(n))/T \to 0$ almost surely as $T \to \infty$, then $n_{r(q)T} \geq n^0_{r(q)}$ with arbitrarily large probability, as $T \to \infty$.

(ii) If $(r(i)_T, n_{r(i)T}) = (r(i)^0, n^0_{r(i)})$, $i = q + 1, \ldots, v$, and $\liminf_{T \to \infty} p_T(d_q(n))/L(T) > 0$ almost surely, where $L(T)$ is a real valued, increasing function of T such that $\log \log T/L(T) \to 0$, then $Pr(\lim_{T \to \infty} n_{r(q)T} \leq n^0_{r(q)}) = 1$.

From Theorem 5.1 it is clear that if $n_{r(j)T} = n_{r(j)0}$ and $r(j)_T = r(j)^0$, for $j = q + 1, \ldots, v$, and provided that $p_T(d_q(n))/T \to 0$ and $\log \log T/p_T(d_q(n)) \to 0$ as $T \to \infty$, then for T sufficiently large we will have $n_{r(q)T} = n^0_{r(q)}$ with probability one. Hence, bar invariant rotations, $r(q)_T$ must coincide with $r(q)^0$ almost surely if $p_T(d_q(n))$ satisfies the requirements of parts (i) and (ii) of Theorem 5.1. Induction on $n_{r(q)T}$ and $r(q)_T$ for $q = 1, \ldots, v$, now yields the following theorem.

Theorem 5.2 *Suppose that $\mathbf{y}_t$ is an ARMA process satisfying Assumptions 1 and 2′, and let $\{r(j)_T, n_{r(j)T}\}$, $j = 1, \ldots, v$, denote the Kronecker invariant pairs obtained by implementing the above algorithm. If $p_T(d_q(n))/T \to 0$ and $\log \log T/p_T(d_q(n)) \to 0$ as $T \to \infty$ then, modulo invariant rotations, $r(j)_T = r(j)^0$ a.s. for T sufficiently large, and $Pr(\lim_{T \to \infty} n_{r(j)T} = n_{r(j)0}) = 1$, $j = 1, \ldots, v$.*

In what follows we will append a zero superscript to quantities of interest to indicate those values corresponding to the actual data generating mechanism giving rise to the observations, as we have already done for the Kronecker invariants. Thus, $\mathbf{\Sigma}^0$ will denote the true system innovation variance-covariance matrix, and $\widetilde{\mathbf{K}}^0(\omega) = \widetilde{\mathbf{A}}^0(\omega)^{-1}\widetilde{\mathbf{M}}^0(\omega)$ will represent the true transfer function of $\widetilde{\mathbf{y}}_t$. Similarly, $\alpha^0(z)$, $\beta^0(z)$ and $\mu^0(z)$ will denote the true autoregressive, exogenous and moving-average operators associated with the scalar ARMAX representation outlined in Proposition 3.2.

6   Adaptations for Cointegrated Processes

Suppose that the operator $\mathbf{A}(z)$ has $v - \varrho$ roots of unity and all other zeroes line outside the unit circle so that $\det\mathbf{A}(z) = a_s(z)(1 - z)^{v-\varrho}$, $a_s(z) \neq 0$, $|z| \leq 1$, $\varrho < v$. Applying the Beveridge-Nelson decomposition $\mathbf{A}(z) = \mathbf{B}(z)(1 - z) + \mathbf{A}(1)z$ in (1.2), where $\mathbf{B}(z) = \mathbf{B}_0 + \mathbf{B}_1 z + \cdots + \mathbf{B}_{p-1}z^{p-1}$ with $\mathbf{B}_0 = \mathbf{A}_0$ and $\mathbf{B}_i = -(\mathbf{A}_{i+1} + \cdots + \mathbf{A}_p)$, $i = 1, \ldots, p - 1$, leads to the (Engle and Granger, 1987) error correction (EC) representation

$$\mathbf{B}(B)\triangle\mathbf{y}_t + \mathbf{C}\mathbf{E}\mathbf{y}_{t-1} = \mathbf{M}(B)\boldsymbol{\varepsilon}_t \,, \qquad (6.1)$$

wherein the cointegrating relations are summarised in the reduced rank representation of the coefficient $\mathbf{A}(1) = \sum_{s=0}^p \mathbf{A}_s = \mathbf{C}\mathbf{E}$ where $\mathbf{C}$ and $\mathbf{E}'$ are $(v \times \varrho)$ matrices with full column rank ρ. Now consider an EC representation in which $[\mathbf{B}(z) : \mathbf{M}(z)]$ satisfy the conditions

(i') $b_{rc,0} = m_{rc,0}$,

(ii) $m_{rr}(z) = 1 + m_{rr,1}z + \ldots + m_{rr,n_r}z^{n_r}$,

   $m_{rc}(z) = m_{rc,n_r-n_{rc}+1}z^{n_r-n_{rc}+1} + \ldots + m_{rc,n_r}z^{n_r}$,

(iii') $b_{rc}(z) = b_{rc,0} + b_{rc,1}z + \ldots + b_{rc,n_r}z^{n_r-1}$,

and the coefficients in the cointegrating relations are normalized such that

(iv') $\mathbf{E}$ is in row-reduced echelon form.

When conditions (i') through (iv') are imposed the structure in (6.1) is canonical and equivalent to an $ARMA_E(\nu)$ representation for $\mathbf{y}_t$ in which the cointegrating rank ϱ has been imposed. A system satisfying these conditions will be labeled an $ECARMA_E(\nu, \varrho)$ form. For detailed particulars see Poskitt (2006).

A row-reduced echelon form is a matrix in which the first nonzero entry in any row is unity and appears to the right of the first nonzero entry in the preceding row, and all other entries in the same column as the first nonzero entry in any row are zero. Imposing this structure on $\mathbf{E}$ provides a solution to the statistical identification problem, but it implies that certain variables can be excluded from the cointegrating relations. Since when identifying the Kronecker invariants we wish to consider different permutations of the variables in $\mathbf{y}_t$ we will work here with a different characterization of the cointegrating space, as different permutations of $\mathbf{y}_t$ may not be compatible with the exclusion constraints implicit in the (admittedly arbitrary) row-reduced echelon form normalization. Following the argument in Poskitt (2000, Remark 1), we can determine a $v \times v$ nonsingular transformation matrix $\mathbf{T} = [\mathbf{T}'_c : \mathbf{T}'_u]'$, where the partition occurs after the first $\varrho \geq 0$ rows, such that $\mathbf{A}(z) = \mathbf{A}_s(z)\mathbf{T}^{-1}\boldsymbol{\Delta}(z)\mathbf{T}$ where $\det\mathbf{A}_s(z) = a_s(z)$ and, without loss of generality, $\boldsymbol{\Delta}(z) = \mathrm{diag}[\mathbf{I}_\varrho : \mathbf{I}_{v-\varrho}(1 - z)]$. Moreover, if $\mathbf{x}_t = [\mathbf{x}'_{ct} : \mathbf{x}'_{ut}]' = \mathbf{T}\mathbf{y}_t$ then $\boldsymbol{\Delta}(B)\mathbf{x}_t = \mathbf{T}\mathbf{A}_s(B)^{-1}\mathbf{M}(B)\boldsymbol{\varepsilon}_t$ and the elements in the first ϱ rows of $\mathbf{x}_t$, the variables in $\mathbf{x}_{ct}$, are asymptotically stationary processes and those in the remaining $v - \varrho$ rows, $\mathbf{x}_{ut}$, are first difference stationary. Note also that $\mathbf{A}(1) = \mathbf{A}_s(1)\mathbf{T}^{-1}\boldsymbol{\Delta}(1)\mathbf{T} = \mathbf{F}\mathbf{T}_c$, say, so $\mathbf{T}_c$ corresponds to the coefficient matrix in a reduced rank factorization of the error correction term.

Proposition 6.1 *Let $\mathbf{y}_t$ be an $ECARMA_E(\nu, \rho)$ process satisfying Assumptions 1 and 2, and suppose that the variables have been ordered according to the Kronecker invariants. Then for each choice of the variable $\triangle z_t = \triangle y_{jt}$, $j = 1, \ldots, v$, there exists coprime polynomial*

operators $\alpha(z)$, $\boldsymbol{\beta}(z)$ and $\mu(z)$ of order $n = n_j$, and a coefficient vector $\mathbf{c}$, such that $\triangle z_t$ admits the representation

$$\alpha(B)\triangle z_t + \boldsymbol{\beta}(B)\triangle\mathbf{y}_t + \mathbf{c}'\mathbf{x}_{c(t-1)} = \mu(B)\eta_t \, , \tag{6.2}$$

wherein $\mathbf{x}_{ct} = \mathbf{T}_c\mathbf{y}_t$, Moreover, in the scalar ARMAX specification (6.2) of $\triangle z_t$ the variables $\triangle y_{it}$, $i = 1, \ldots, j-1$, $\triangle y_{it-s}$, $i = 1, \ldots, v$, $i \neq j$, $s = 1, \ldots, n$, and $\mathbf{x}_{c(t-1)}$, are stationary predetermined regressors.

7   Identification Algorithm: Cointegrated Case

The following algorithm uses the structure in Proposition 6.1 to identify the cointegrating rank and the Kronecker invariants in two stages. The first stage identifies ρ and a basis for $\mathbf{T}_c$, a basis for the cointegrating space. The second stage then exploits the formulation in (6.2) and identifies the Kronecker invariants using a modification of the previous algorithm.

ALGORITHM–$ECARMA_E(\nu, \varrho)$.
Stage 1: Determine the solutions to the eigenvalue-eigenvector problem

$$[\lambda\mathbf{I}_v - \mathbf{C}_y(0)^{-1}\mathbf{C}_y(1)\mathbf{C}_y(0)^{-1}\mathbf{C}_y(1)']\mathbf{v}_T = \mathbf{0}$$

where the pairs $[\lambda_{(i),T} : \mathbf{v}_{(i),T}]$, $i = 1, \ldots, v$, are arranged according to the ordering $\lambda_{(1),T} \leq \lambda_{(2),T} \leq \cdots \leq \lambda_{(v),T}$ of the eigenvalues and the associated eigenvectors are normalised such that $\mathbf{v}'_{(i),T}\mathbf{C}_y(0)\mathbf{v}_{(j),T} = \delta_{i,j}$, the Kronecker delta. For $\varrho = 0, \ldots, v-1$ determine

$$\nabla_T(\varrho) = T(v - \varrho)\log\left(\sum_{j=\varrho+1}^{v}\lambda_{(j),T}/(v-\varrho)\right) - T\sum_{j=\varrho+1}^{v}\log(\lambda_{(j),T})$$
$$+ \varrho(2v - \varrho + 1)\log\log T/2$$

and set

$$\varrho = \min_{0\leq\varrho<v}\nabla_T(\varrho)\,.$$

Now set $\mathbf{B}_T = [\mathbf{v}_{(1),T} : \cdots : \mathbf{v}_{(\varrho),T}]'$.
   Stage 2:

Initialization: Set $n = 1$, $j = v$, $\mathbf{P} = \mathbf{I}$ and $\mathcal{N} = \{1, \ldots, v\}$. Compute the first differenced values $\triangle\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ and the regressor $\boldsymbol{\zeta}_t = \mathbf{B}_T\mathbf{y}_t$ for $t = 1, \ldots, T$. For each $i \in \mathcal{N}$, set $\widetilde{\sigma}^2_{\eta,i}(0)$ equal to the residual mean square from the regression of $\triangle y_{it}$ on $\triangle y_{kt}$, $k = 1, \ldots, v$, $k \neq i$, and $\boldsymbol{\zeta}_{(t-1)}$.

while: $j \geq 1$

   for $i(k) \in \mathcal{N}$, $k = 1, \ldots, v$,
      1. Set $\triangle\widetilde{\mathbf{y}}_t = \mathbf{E}_{i(k),j}[\mathbf{P}\triangle\mathbf{y}_t]$ and for $\triangle z_t = \triangle y_{i(k)t}$ evaluate initial estimates of the jth scalar ARMAX form :

(a) Construct estimates of the nonzero coefficients in $\widetilde{\mathbf{d}}(z) = \mathbf{e}_j'\widetilde{\mathbf{D}}(z)$, the jth row of $\widetilde{\mathbf{D}}(z) = [\mathbf{E}_{i(k),j}\mathbf{P}\mathbf{B}(z)\mathbf{P}'\mathbf{E}'_{i(k),j}, \mathbf{E}_{i(k),j}\mathbf{P}\mathbf{C}]$, by solving the equations

$$\int_{-\pi}^{\pi} \widehat{\widetilde{\mathbf{d}}}(\omega)\widetilde{\mathbf{I}}_w(\omega)\exp(-\imath\omega r)d\omega = \mathbf{0}, \ r = n+1,\ldots,2n, \qquad (7.1)$$

for $\widetilde{\mathbf{d}}(\omega) = \widetilde{\mathbf{d}}_0 + \widetilde{\mathbf{d}}_1\exp(\imath\omega) + \cdots + \widetilde{\mathbf{d}}_n\exp(\imath\omega n) = \widetilde{\mathbf{d}}(z)\big|_{z=e^{\imath\omega}}$ where

$$\widetilde{\mathbf{I}}_w(\omega) = \frac{1}{2\pi}\sum_{s=-T+1}^{T-1}\widetilde{\mathbf{C}}_w(s)\exp(\imath\omega)$$

and

$$\widetilde{\mathbf{C}}_w(r) = T^{-1}\sum_{t=1}^{T-|r|}\widetilde{\mathbf{w}}_t\widetilde{\mathbf{w}}'_{t-r}$$

for $r = 1,\cdots,T-1$ where $\widetilde{\mathbf{w}}_t = [\triangle\widetilde{\mathbf{y}}'_t, \boldsymbol{\zeta}'_{t-1}]'$. Now set $\widehat{\widetilde{\alpha}}_s = \widehat{\widetilde{d}}_{j,s}$, $s = 1,\ldots,n$, $\widehat{\widetilde{\beta}}_i = \widehat{\widetilde{d}}_{i,0}$, $i = 1,\ldots,j-1$, $\widehat{\widetilde{\beta}}_{i,s} = \widehat{\widetilde{d}}_{i,s}$, $i = 1,\ldots,v$, $i \neq j$, $s = 1,\ldots,n$ and $c_r = \widehat{\widetilde{d}}_{v+r,1}$, $r = 1,\ldots,\varrho$.

(b) For $r = 1,\cdots,n$ form

$$\widetilde{C}_v(r) = \int_{-\pi}^{\pi}\widehat{\widetilde{\mathbf{d}}}(\omega)\widetilde{\mathbf{I}}_w(\omega)\widehat{\widetilde{\mathbf{d}}}(\omega)^*\exp(-\imath\omega r)d\omega$$

and set

$$\widehat{\widetilde{S}}_v(\omega) = \frac{1}{2\pi}\sum_{s=-n}^{n}\widetilde{C}_v(s)\exp(\imath\omega s). \qquad (7.2)$$

Compute estimates of the moving average coefficients $\widehat{\widetilde{\mu}}_s$, $s = 1,\ldots,n$, by solving the equation system

$$\sum_{s=0}^{n}\widehat{\widetilde{\mu}}_s\widetilde{C}i_v(l+s) = 0, \ l = 1,\ldots,n$$

where

$$\widetilde{C}i_v(r) = \int_{-\pi}^{\pi}\frac{\widehat{\widetilde{\mathbf{d}}}(\omega)\widetilde{\mathbf{I}}_w(\omega)\widehat{\widetilde{\mathbf{d}}}(\omega)^*}{\widehat{\widetilde{S}}_v(\omega)^2}\exp(-\imath\omega r)d\omega. \qquad (7.3)$$

2. Compute a pseudo maximum likelihood estimate (pseudo MLE) of the innovation variance $\widetilde{\sigma}_\eta^2$.

(a) Form the $v + \varrho + 1$ vector sequence $(\widehat{\widetilde{\boldsymbol{\xi}}}'_t, \widehat{\widetilde{\varphi}}_t,)'$ by solving

$$\begin{bmatrix} \widehat{\widetilde{\boldsymbol{\xi}}}_t \\ \widehat{\widetilde{\varphi}}_t \end{bmatrix} + \sum_{s=1}^{n}\widehat{\widetilde{\mu}}_s\begin{bmatrix} \widehat{\widetilde{\boldsymbol{\xi}}}_{t-s} \\ \widehat{\widetilde{\varphi}}_{t-s} \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{w}}_t \\ \widehat{\widetilde{\eta}}_t \end{bmatrix}$$

for $t = 1,\ldots,T$ where

$$\widehat{\widetilde{\eta}}_t = \widehat{\widetilde{\mathbf{d}}}(B)\widetilde{\mathbf{w}}_t + (1 - \widehat{\widetilde{\mu}}(B))\widehat{\widetilde{\eta}}_t$$

and the recursions are initiated at $\widehat{\widetilde{\eta}}_t = 0$ and $(\widehat{\widetilde{\boldsymbol{\xi}}}'_t, \widehat{\widetilde{\varphi}}_t,)' = \mathbf{0}'$, $t \leq 0$.

(b) Set $\widetilde{c}_{\widehat{\eta}}(0) = T^{-1} \sum_t^T \widehat{\widetilde{\eta}}_t^2$ and calculate

$$\widetilde{\mathbf{c}}_{\widehat{\eta}\widehat{\xi}}(s) = T^{-1} \sum_{t=s+1}^{T} \widehat{\widetilde{\eta}}_t \widehat{\widetilde{\boldsymbol{\xi}}}_{t-s}' \quad \text{and} \quad \widetilde{c}_{\widehat{\eta}\widehat{\varphi}}(s) = T^{-1} \sum_{t=s+1}^{T} \widehat{\widetilde{\eta}}_t \widehat{\widetilde{\varphi}}_{t-s}.$$

Now compute the mean squared error

$$\widehat{\widetilde{\sigma}}_\eta^2(n) = \widetilde{c}_{\widehat{\eta}}(0) + \sum_{s=0}^{n} \Delta\widehat{\widetilde{\mathbf{d}}}_s \widetilde{\mathbf{c}}_{\widehat{\eta}\widehat{\xi}}(s)' - \sum_{s=1}^{n} \Delta\widehat{\widetilde{\mu}}_s \widetilde{c}_{\widehat{\eta}\widehat{\varphi}}(s)$$

where $\Delta\widehat{\widetilde{\mathbf{d}}}_s$, $s = 0, \ldots, n$, and so on, denote the coefficient values obtained from the Toeplitz regression of $\widehat{\widetilde{\eta}}_t$ on $-\widehat{\widetilde{\xi}}_{it}$, $i = 1, \ldots, j-1$, and $-\widehat{\widetilde{\xi}}_{(v+r)t}$, $r = 1, \ldots, \varrho$, and $-\widehat{\widetilde{\boldsymbol{\xi}}}_{t-s}$ and $\widehat{\widetilde{\varphi}}_{t-s}$, $s = 1, \ldots, n$.

3. Apply model selection rule:

   (a) Evaluate the criterion function

   $$IC_T(n) = T \log \widehat{\widetilde{\sigma}}_\eta^2(n) + p_T(d_j(n))$$

   where the penalty term $p_T(d_j(n)) > 0$ is a real valued function, monotonically increasing in $d_j(n)$ and non-decreasing in T.

   (b) **if** $IC_T(n) > IC_T(n-1)$;

       set $r(j) = i(k)$; $n_{r(j)} = n - 1$; and

       update $\mathbf{P} = \mathbf{E}_{i(k),j}\mathbf{P}$; $\mathcal{N} = \mathcal{N} \setminus i(k)$; $j = j - 1$.

       end

end for $i(k) \ni \mathcal{N}$.

  if $n < h_T$;

    increment $n = n + 1$.

  else

    for $i(k) \in \mathcal{N}$, $k = 1, \ldots, v$,

      set $n_{r(j)} = n$; $r(j) = i(k)$; and

      update $\mathbf{P} = \mathbf{E}_{i(k),j}\mathbf{P}$; $\mathcal{N} = \mathcal{N} \setminus i(k)$; $j = j - 1$.

      end for $i(k) \ni \mathcal{N}$.

  end

end when $j = 0$.

REMARK 6. The eigenvalues and eigenvectors calculated in the first stage of Algorithm–$ECARMA_E(\nu, \varrho)$ correspond to the *raw* canonical correlations and discriminant functions (using classical nomenclature) between $\mathbf{y}_t$ and $\mathbf{y}_{t-1}$ (See Box and Tiao, 1977; Poskitt, 2000). The estimate of the cointegrating rank, ϱ_T, and the basis for the cointegrating space, $\mathbf{B}_T$, are known functions of these canonical variables. Unlike the Johansen procedure (Johansen, 1995), the technique employed here for identifying the cointegrating structure does not require the specification of an underlying parametric model, both ϱ_T and $\mathbf{B}_T$ are nonparametric in nature and can be determined without knowledge of the Kronecker invariants.

8   Further Theoretical Properties

Theorem 8.1 below is derived from a sequence of lemmas that recognize that since by Theorem 2.2 of Poskitt (2000) ϱ_T provides a strongly consistent estimate of the cointegrating rank, $\varrho_T = \varrho^0$ with probability one for T sufficiently large and we can therefore assume for the purposes of asymptotic derivations that $\varrho_T = \varrho^0$. Given $\varrho_T = \varrho^0$, Lemma 3.1 of Poskitt (2000) indicates that $\mathbf{B}_T$ provides a super-strongly consistent estimate of a basis for the cointegrating space in that there exists a nonsingular matrix $\mathbf{R}_T$ such that $T[\mathbf{B}_T - \mathbf{R}_T \mathbf{T}_c^0] = O(1)$ as $T \to \infty$. Now, noting that $\mathbf{F} \mathbf{R}_T^{-1} \boldsymbol{\zeta}_{t-1} - \mathbf{F} \mathbf{x}_{c(t-1)} = \mathbf{F} \mathbf{R}_T^{-1} (\mathbf{B}_T - \mathbf{R}_T \mathbf{T}_c^0) \mathbf{y}_{t-1}$, it follows that the regressors $\boldsymbol{\zeta}_{t-1}$ and $\mathbf{x}_{c(t-1)}$ are asymptotically equivalent up to a nonsingular transformation determined by $\mathbf{R}_T$. As is shown in the Appendix, a consequence of this is that the value $\widehat{\widetilde{\sigma}}_\eta^2(n)$ obtained at the second stage of Algorithm–$ARMA_E(\nu, \varrho)$ is asymptotically equivalent to the value that would have been obtained had $\boldsymbol{\zeta}_{t-1}$ been replaced by $\mathbf{x}_{c(t-1)}$. This equivalence, namely, that for T sufficiently large the mean squared error estimates based on $\triangle \mathbf{y}_{t-j}$, $j = 0, \ldots, n$, and $\boldsymbol{\zeta}_{t-1}$, and $\triangle \mathbf{y}_{t-j}$, $j = 0, \ldots, n$, and $\mathbf{x}_{c(t-1)}$, are equal almost surely, underpins the following result.

Theorem 8.1 *Suppose that $\mathbf{y}_t$ is an $ECARMA_E(\nu, \rho)$ process satisfying Assumptions 1 and 2, and that $\varrho_T = \varrho^0$. Let $\{r(j)_T, n_{r(j)T}\}$, $j = 1, \ldots, v$, denote the Kronecker invariant pairs obtained when employing Algorithm $ECARMA_E(\nu, \varrho)$ with $p_T(d_j(n))$ a possibly stochastic function of n and T. Then:*

(i)   If $(r(i)_T, n_{r(i)T}) = (r(i)^0, n_{r(i)}^0)$, $i = q + 1, \ldots, v$, and $p_T(d_q(n))/T \to 0$ almost surely as $T \to \infty$, then $n_{r(q)T} \geq n_{r(q)}^0$ with arbitrarily large probability, as $T \to \infty$.

(ii)   If $(r(i)_T, n_{r(i)T}) = (r(i)^0, n_{r(i)}^0)$, $i = q + 1, \ldots, v$, and $liminf_{T\to\infty} p_T(d_q(n))/L(T) > 0$ almost surely, where $L(T)$ is a real valued, increasing function of T such that $loglog T/L(T) \to 0$, then $Pr(\lim_{T\to\infty} n_{r(q)T} \leq n_{r(q)}^0) = 1$.

The first part of the following theorem is simply a restatement of the fact that $Pr(\lim_{T\to\infty} \varrho_T = \varrho^0) = 1$ (Poskitt, 2000, Theorem 2.2).

Theorem 8.2 *Suppose that $\mathbf{y}_t$ is an $ECARMA_E(\nu, \rho)$ process satisfying Assumptions 1 and 2. Let ϱ_T and $\{r(j)_T, n_{r(j)T}\}$, $j = 1, \ldots, v$, denote the cointegrating rank and Kronecker invariant pairs obtained by implementing the above algorithm. Then for T sufficiently large $\varrho_T = \varrho^0$ with probability one, and if $p_T(d_q(n))/T \to 0$ and $\log\log T/p_T(d_q(n)) \to 0$ as $T \to \infty$ then, modulo invariant rotations, $r(j)_T = r(j)^0$ a.s. for T sufficiently large, and $Pr(\lim_{T\to\infty} n_{r(j)T} = n_{r(j)0}) = 1$, $j = 1, \ldots, v$.*

The latter part of Theorem 8.2 can be deduced from Theorem 8.1 in the same way that Theorem 5.2 follows on from Theorem 5.1.

9   Illustration and Conclusion

In this paper a new methodology for identifying the structure of VARMA models has been developed. The approach that has been adopted is to construct an estimate of the echelon

canonical form using a new technique to determine the Kronecker invariants. Algorithms that are applicable to both asymptotically stationary and cointegrated time series were presented. The algorithms facilitate a fully automatic approach to model specification, and the estimates that are so provided have been shown to be strongly consistent.

A novel feature of the inferential procedures developed here is that they work in terms of a canonical scalar ARMAX representation for each variable in which the exogenous regressors are chosen from predetermined contemporaneous and lagged values of other variables in the VARMA system. By working in terms of the scalar ARMAX specification the algorithms address two issues. First, using the scalar ARMAX specification reduces the range of VARMA structures that needs to be investigated by several orders of magnitude, significantly ameliorating the curse of dimensionality inherent in the analysis of VARMA models. Second, any problems that might be inherent in using a (long) VAR in a macroeconomic modeling context are avoided, since use of the scalar ARMAX specification allows parameter estimates to be constructed directly, obviating the need to use autoregressive approximations.

To illustrate the inferential potential of the methods advanced here, Figure 1 graphs
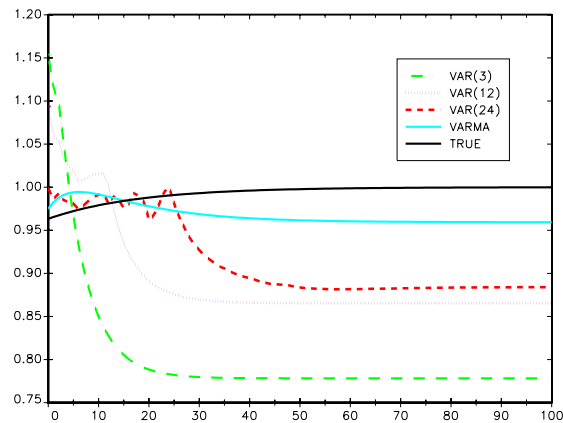
(a) Impulse Response of $\log Y_t$
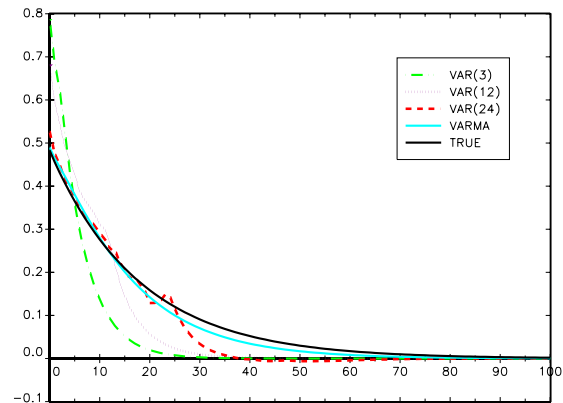
(b) Impulse Response of n_t

Figure 1: RBC Model: Impulse Responses to a Technology Shock

the theoretical impulse responses derived from a real business cycle (RBC) model together with those calculated from four statistical models fitted to data generated from the model. The experiment uses the RBC model due to Hansen (1985), and following Ravenna (2007)

and Erceg, Guerrieri and Gust (2005) takes the capital stock as unobservable and treats total output, Y_t, and hours worked, n_t, as the observable variables available for empirical modelling. The model was calibrated as in Ravenna (2007) and Erceg et al. (2005) and it can be shown (Yao and Kam, 2011) that the parameter values imply that the data generating mechanism for $\mathbf{y}_t = (n_t, \Delta \log Y_t)'$ has the representation

$$\begin{pmatrix} 1 - 0.941B & -1.045B \\ -0.770 + 0.724B & 1 \end{pmatrix} \mathbf{y}_t = \begin{pmatrix} 1 - 0.250B & -0.917B \\ -0.770 & 1 \end{pmatrix} \boldsymbol{\varepsilon}_t, \qquad (9.1)$$

a bivariate $ARMA_E(\{1, 1\})$ process where the elements of the innovation $\boldsymbol{\varepsilon}_t$ equal linear combinations of two uncorrelated exogenous structural shocks, a technology shock and a labor supply shock. The four statistical models comprise an estimated VARMA model and three VAR approximations constructed from a realization of 20000 observations on $\mathbf{y}_t$.

For each model the transformation from the innovations to the structural shocks is obtained using the Blanchard and Quah (1989) identification strategy. Neither structural shock has a permanent impact on either component of $\mathbf{y}_t$, since $\mathbf{y}_t$ is stationary, and the long run identifying restrictions imply that a labor supply shock has no long run effect on the level of Y_t, while the opposite is true for a technology shock.

Preliminary analysis of the raw simulated data n_t and Y_t, $t = 1, \ldots, T = 20000$, indicated that $\mathbf{y}_t = (n_t, \Delta \log Y_t)'$ could be treated as a stationary process. An augmented Dickey–Fuller test unequivocally rejected the hypothesis that n_t was a unit root process at all conventional significance levels, whereas the same was not true of $\log Y_t$, and ϱ_T unambiguously indicated that n_t and $\log Y_t$ have a cointegrating rank of one. The VARMA model for $\mathbf{y}_t$ was correctly identified as an $ARMA_E(\{1, 1\})$, and eliminating redundant parameters using the rules outlined in Athanasopoulos and Vahid (2008) gave the following estimated structure

$$\begin{pmatrix} 1 - 0.941B & -1.003B \\ -0.785 + 0.740B & 1 - 0.005B \end{pmatrix} \mathbf{y}_t = \begin{pmatrix} 1 - 0.257B & -0.870B \\ -0.785 & 1 \end{pmatrix} \widehat{\boldsymbol{\varepsilon}}_t \qquad (9.2)$$

when the resulting model was fitted using FIML. The VAR approximations were estimated using (Gaussian) QML with different autoregressive orders, h, chosen by reference to AIC, BIC and HQ, giving $h = 24, 3$ and 12 respectively.

Given the proximity of the parameter values in (9.2) to those in (9.1) it is perhaps not surprising to find that the VARMA modelling procedure has produced reasonable estimates of the impulse response functions, and it is of interest to note that the impulse responses generated from an unconstrained $ARMA_E(\{1, 1\})$ model are almost indistinguishable from those generated from (9.2). The worst fitting impulse response profile from the VARMA model is that estimating the impact of a technology shock on total output, but even here the long run effect is under estimated by less than 5%.

It is clear from both panels in Figure 1, however, that the autoregressive approximations do not mimic the true dynamics of the theoretical model very well, even with the longest lag length. Interestingly enough, increasing the value of h does little to improve the general shape of the VAR impulse response estimates, it merely serves to increase the minor oscillations

around the overall impulse response profiles given by the VAR(24) approximation. This feature is clearly seen in Figure 2. Figure 2 graphs the theoretical impulse responses together with those calculated from the estimated VARMA model, the VAR(24) and three additional VAR(h) approximations with $h = 30, 50$ and 100, corresponding to $h = (\log T)^a$ with $a = 1.48, 1.7$ and 2.0 respectively.
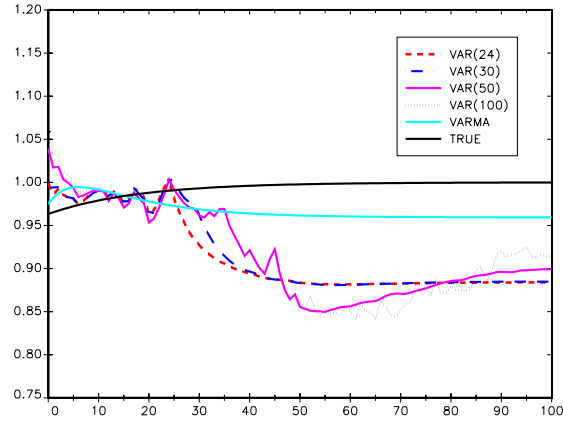
(a) Impulse Response of $\log Y_t$
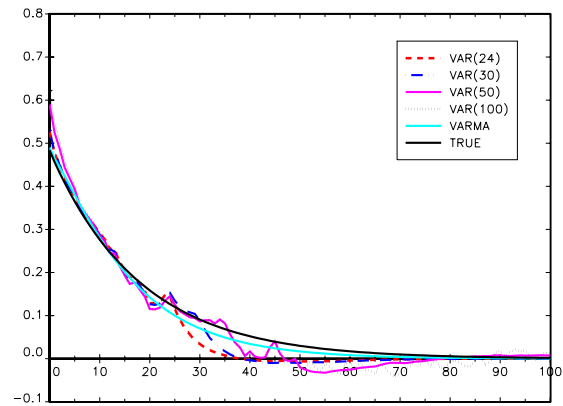
(b) Impulse Response of n_t

Figure 2: RBC Model: Impulse Responses to a Technology Shock

These large sample outcomes amply demonstrate the concepts, ideas and results discussed above and lend strong support to the conclusions of Kapetanios et al. (2007). When dealing with real macroeconomic data, however, it is usually the case that only limited sample sizes are available, typically around 200 with quarterly observations, and hence a comparison between VARMA models and VAR approximations based on data from small samples may be of more practical importance.

At this smaller sample size the structure inherent in the RBC model is not reliably estimated. All three criteria, AIC, BIC and HQ, select an autoregression with $h = 1$. This is related to the fact that the zeroes of $\mathbf{M}(z)$ in (9.1) are 1.0459 and ∞, and those of $\mathbf{A}(z)$ are 1.0584 and 1.2488. The proximity of the two smallest roots suggests that a near cancelation occurs, a phenomenon previously encountered by Cogley and Nason (1993). As a consequence, the coefficients in the VARMA model cannot be estimated with any degree of precision unless T is very large, and this results in the impulse responses calculated from the

estimated VARMA model tracking their VAR(1) approximation counterparts fairly closely when $T = 200$, the performance of both being equally disappointing. Changing the parameter settings in the Hansen (1985) RBC model does little to alter these small sample properties and highlights a challenging problem: Even though a theoretical model might lead to a VARMA data generating mechanism, with insufficient observations it may be hard to distinguish inferences based on VARMA models from those produced using VAR approximations. Such outcomes are in accord with the findings of Kascha and Mertens (2009), who suggest that the disadvantages of using VAR approximations rather than VARMA models may therefore be over stated.

Recent work of Yao and Kam (2011) indicates, however, that behaviour of the type seen above when $T = 20000$ can also be observed at sample sizes as small as $T = 200$. They consider a two sector business–cycle monetary–search model that is an adaptation of Hansen (1985) along the lines of Lagos and Wright (2005). Their model nests the one used here and introduces more complicated interactions that result in the observed variables having a more complex stochastic structure. Yao and Kam (2011) confirm that with large samples the true $ARMA_E(\nu)$ form is identified almost surely and that the VARMA modelling procedure can produce quite precise impulse response function estimates. Their results also show that VARs can yield poor approximations in spite of T being large (*c.f.* Chari et al., 2007). Furthermore, Yao and Kam (2011) find that even at small sample sizes VARMA models yield better estimates of the dynamics of their true business–cycle monetary–search model than VAR approximations.

In summary, existing experimental evidence on the relative merits of using VARMA models vis–â–vis VAR approximations for macroeconomic modelling indicates that performance is critically influenced by the sample size examined and, perhaps less obviously but possibly more importantly, by the choice of economic model underlying the simulated data generating mechanism. When taken together the overall conclusions are at best inconclusive. Although asymptotic properties and results of the type presented in this paper can give an indication of the likely behaviour of the procedures, their scope may be limited to very large samples and additional insight into performance in small finite sample situations is needed. It is hoped to present some guidance on the latter via more extensive simulation experiments and empirical applications elsewhere.

References

AKAIKE, H. (1974a). A new look at the statistical model identification. *IEEE-Transactions on Automatic Control* **AC-19** 716–723.

AKAIKE, H. (1974b). Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control* **AC-19** 667–674.

ATHANASOPOULOS, G., POSKITT, D. S. and VAHID, F. (2010). Two canonical VARMA forms: Scalar component models vis-à-vis the Echelon form. *Econometric Reviews* In Press.

ATHANASOPOULOS, G. and VAHID, F. (2008). A complete VARMA modelling methodology based on scalar components. *Journal of Time Series Analysis* **29** 533–554.

BARTEL, H. and LÜTKEPOHL, H. (1998). Estimating the kronecker indices of cointegrated Echelon-form VARMA models. *Econometrics Journal* **1** 76–99.

BINGHAM, E. O. (1974). *The Fast Fourier Transform.* Prentice Hall, Englewood Cliffs.

BLANCHARD, O. J. and QUAH, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review* **79** 655–73.

BOX, G. E. P. and TIAO, G. C. (1977). A canonical analysis of multiple time series. *Biometrika* **64** 355–365.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rienhart and Winston, New York.

CHARI, V. V., KEHOE, P. J. and MCGRATTAN, E. R. (2007). Are structural VARs with long-run restrictions useful in developing business cycle theory? *Federal Reserve Bank of Minneapolis, Research Department Staff Report 364* .

CHRISTIANO, L. J., EICHENBAUM, M. and VIGFUSSON, R. (2006). Assessing structural VARs. *NBER Macroeconomics Annual* 1–106.

COGLEY, T. and NASON, J. M. (1993). Impulse dynamics and propagation mechanisms in a real business cycle model. *Economics Letters* **43** 77–81.

COOLEY, T. and DWYER, M. (1998). Business cycle analysis without much theory. A look at structural VARs. *Journal of Econometrics* **83** 57–88.

DAHLHAUS, R. (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Annals of Statistics* **16** 808–841.

DEL NEGRO, M. and SCHRFHEIDE, F. (2004). Priors from general equilibrium models for VARs. *International Economic Review* **45** 643–673.

DOAN, T., LITTERMAN, R. and SIMS, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* **3** 1–10.

ENGLE, R. F. and GRANGER, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55** 251–276.

ERCEG, C. J., GUERRIERI, L. and GUST, C. (2005). Can long-run restrictions identify technology shocks? *Journal of the European Economic Association* **3** 1237–1278.

FERNÁNDEZ-VILLAVERDE, J., RUBIO-RAMÍREZ, J. F., SARGENT, T. J. and WATSON, M. W. (2007). ABC's (and D)'s of understanding VARs. *The American Economic Review* **97** 1021–1026.

FRY, R. and PAGAN, A. (2005). Some issues in using VARs for macroeconometric research. *CAMA Working Paper Series 19/2005, Australian National University* .

HANNAN, E. J. (1971). *Multiple Time Series.* J. Wiley, New York.

HANNAN, E. J. and DEISTLER, M. (1988). *The Statistical Theory of Linear Systems.* Wiley, New York.

HANNAN, E. J. and KAVALIERIS, L. (1984). Multivariate linear time series models. *Advances in Applied Probability* **16** 492–561.

HANNAN, E. J. and QUIN, B. G. (1979). The determination of the order of an autoregression. *Journal of Royal Statistical Society* **B 41** 190–195.

HANNAN, E. J. and RISSANEN, J. (1982). Recursive estimation of autoregressive-moving average order. *Biometrica* **69** 81–94.

HANSEN, G. D. (1985). Indivisible labor and the business cycle. *Journal of Monetary Economics* **16** 309–327.

JOHANSEN, S. (1995). *Likelihood-based Inference on Cointegrated Vector Auto-regressive Models.* Oxford University Press, Oxford.

KAPETANIOS, G., PAGAN, A. and SCOTT, A. (2007). Making a match: combining theory and evidence in policy-oriented macroeconomic modelling. *Journal of Econometrics* **136** 565–594.

KASCHA, C. and MERTENS, K. (2009). Business cycle analysis and varma models. *Journal of Economic Dynamics and Control* **33** 267 – 282.

LAGOS, R. and WRIGHT, R. (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy* **113** 463–484.

LÜTKEPOHL, H. (2005). *The New Introduction to Multiple Time Series Analysis.* Springer-Verlag, Berlin.

LÜTKEPOHL, H. and CLAESSEN, H. (1997). Analysis of cointegrated VARMA processes. *Journal of Econometrics* **80** 223–239.

NSIRI, S. and ROY, R. (1992). On the identification of ARMA echelon form models. *Canadian Journal of Statistics* **20** 369–386.

PAULSEN, J. and TJØSTHEIM, D. (1985). On the estimation of residual variance and order in autoregressive time series. *Journal of the Royal Statistical Society* **B-47** 216–228.

POSKITT, D. S. (1992). Identification of Echelon canonical forms for vector linear processes using least squares. *The Annals of Statistics* **20** 195–215.

POSKITT, D. S. (2000). Strongly consistent determination of cointegrating rank via canonical correlations. *Journal of Business and Economic Statistics* **18** 71–90.

POSKITT, D. S. (2003). On the specification of cointegrated autoregressive moving-average forecasting systems. *International Journal of Forecasting* **19** 503–519.

POSKITT, D. S. (2005). A note on the specification and estimation of ARMAX systems. *Journal of Time Series Analysis* **26** 157–183.

POSKITT, D. S. (2006). On the identification and estimation of nonstationary and cointegrated ARMAX systems. *Econometric Theory* **22** 1138–1175.

POSKITT, D. S. and CHUNG, S. (1996). Markov chain models, time series analysis and extreme value theory. *Advances in Applied Probability* **28** 405–425.

PÖTSCHER, B. M. (1983). Order estimation in arma-models by lagrangian multiplier tests. *Annals of Statistics* **11** 872–885.

RAVENNA, F. (2007). Vector autoregressions and reduced form representations of DSGE models. *Journal of Monetary Economics* **54** 2048–2064.

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

SIMS, C. A. (1980). Macroeconomics and reality. *Econometrica* **48** 1–48.

SMITH, A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8** 563–584.

STOICA, P., SODERSTROM, T. and FRIEDLANDER, B. (1985). Optimal instrumental variables estimates of the AR parameters of an ARMA process. *IEEE Transactions on Automatic Control* **AC–30** 1066–1074.

TIAO, G. C. and TSAY, R. S. (1989). Model specification in multivariate time series (with discussions). *Journal of the Royal Statistical Society B* **51** 157–213.

TJØSTHEIM, D. and PAULSEN, J. (1983). Bias of some commonly-used time series estimates. *Biometrika* **70** 389–399.

WALLIS, K. (1977). Multiple time series analysis and the final form of econometric models. *Econometrica* **45** 1481–1497.

YAO, W. and KAM, T. (2011). Reassessing business-cycle VAR and VARMA models. Monash University, Draft Working Paper.

Appendix: Proofs

Proof of Lemma 3.1:   That $\mathbf{u}_t$ is a moving-average process of order p is obvious from expression (1.2). Now let $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)'$ denote the j'th v element Euclidean basis vector. From standard results on linear filtering we know that the spectral density of $\nu_t = \mathbf{e}_j' \mathbf{u}_t$ is

$$S_\nu(\omega) = \frac{1}{2\pi} \mathbf{e}_j' \mathbf{M}(\omega) \boldsymbol{\Sigma} \mathbf{M}(\omega)^* \mathbf{e}_j \,,$$

with corresponding autocovariance generating function

$$
\begin{aligned}
\rho(z) &= \sum_{s=-n}^{n} \rho_s z^s \\
&= \mathbf{e}_j' \mathbf{M}(z) \boldsymbol{\Sigma} \mathbf{M}(z^{-1})' \mathbf{e}_j \,,
\end{aligned}
$$

where $n = n_j$. The remainder of the proof is standard. The polynomial $z^n \rho(z)$ has $2n$ roots and since the coefficients $\rho_s = \rho_{-s}$, $s = 1, \ldots, n$, of $\rho(z)$ are real, the roots are real or occur in complex conjugate pairs. Now, $\rho(\omega) = 2\pi S_\nu(\omega) > 0$, $-\pi < \omega \le \pi$, so $\rho(z)$ has no zeroes on the unit circle and we may number and group the roots into two sets $\{\zeta_1, \ldots, \zeta_n\}$ and $\{\bar{\zeta}_1^{-1}, \ldots, \bar{\zeta}_n^{-1}\}$ such that $|\zeta_s| < 1$, $s = 1, \ldots, n$. Hence we can construct the operator

$$m(z) = m_0 \prod_{s=1}^{n} (1 - \zeta_s z)$$

where $m_0 = \{\prod (-\zeta_s) \rho_n\}^{\frac{1}{2}}$ such that

$$
\begin{aligned}
\rho(z) &= \rho_n z^{-n} \prod_{s=1}^{n} (1 - \zeta_s z)(1 - \bar{\zeta}_s^{-1} z) \\
&= m(z) m(z^{-1}) \,.
\end{aligned}
$$

Thus we can select n roots of $\rho(z)$ to construct $\mu(z) = 1 + \mu_1 z + \ldots + \mu_n z^n$ such that $\rho(z) = \sigma_\eta^2 \mu(z) \mu(z^{-1})$, where $\mu(z) = m(z)/m_0$ and $\sigma_\eta^2 = m_0^2$, and the roots are chosen in such a way that $\mu(z) \ne 0$, $|z| \le 1$. The existence of the white noise process η_t providing a moving-average representation of ν_t now follows from the spectral factorization theorem, Rozanov (1967, Theorem 9.1, p. 41) (*c.f.* Lütkepohl, 2005, §11.6). $\square$

Proof of Proposition 3.1:   To begin, recall that for $\mathbf{y}_t$ ordered according to the Kronecker invariants the echelon form in (2.1) is such that $\mathbf{A}_0 = \mathbf{M}_0$ is lower triangular and the system representation is (contemporaneously) recursive. Let $\mathbf{a}(z) = \mathbf{e}_j' \mathbf{A}(z)$ denote the nonzero coefficients in the jth row of $\mathbf{A}(z)$. Then, for the jth equation of (1.2) we have

$$\mathbf{a}(B)\mathbf{y}_t = y_{jt} + \sum_{i=1}^{j-1} a_{ji,0} y_{it} + \sum_{i=1}^{v} \sum_{s=1}^{n_j} a_{ji,s} y_{it-s} = u_{jt} \tag{A.1}$$

where, by Lemma 3.1,

$$u_{jt} = \nu_t = \eta_t + \mu_1 \eta_{t-1} + \ldots + \mu_n \eta_{t-n} \,. \tag{A.2}$$

Setting $z_t = y_{jt}$ and reorganizing (A.1), adopting the notational conventions $\alpha_s = a_{jj,s}$, $s = 1, \ldots, n = n_j$, $\beta_i = a_{ji,0}$, $i = 1, \ldots, j-1$, and $\beta_{i,s} = a_{ji,s}$, $i = 1, \ldots, v$, $i \neq j$, $s = 1, \ldots, n$, now gives us the scalar ARMAX specification in (3.2).

To show that $\mathbf{a}(z)$ and $\mu(z)$, and hence the polynomials $\alpha(z)$, $\boldsymbol{\beta}(z)$ and $\mu(z)$, are coprime, assume otherwise. Then we could cancel the common factors in the representation $\mathbf{a}(B)\mathbf{y}_t = \mu(B)\eta_t$ obtained by combining (A.1) and (A.2) to give $\bar{a}(B)\mathbf{y}_t = \bar{\mu}(B)\eta_t$ where $\delta[\bar{\mathbf{a}}, \bar{\mu}] < n_j$. Implementing the same technique as employed in Poskitt (2005, pp. 179-180), we could now use rows $1, \ldots, j-1$ and $j+1, \ldots, v$ of (1.2), together with $\bar{a}(B)\mathbf{y}_t = \bar{\mu}(B)\eta_t$, to construct a unimodular matrix $\bar{\mathbf{U}}(z) \neq \mathbf{I}$ and an operator pair $[\bar{\mathbf{A}}(z) : \bar{\mathbf{M}}(z)] \in ARMA_E(\bar{\nu})$, with multi-index $\bar{\nu} = \{\bar{n}_1, \ldots, \bar{n}_v\}$, $\bar{n}_j \leq n_j$, such that $[\mathbf{A}(z) : \mathbf{M}(z)] = \bar{\mathbf{U}}(z)[\bar{\mathbf{A}}(z) : \bar{\mathbf{M}}(z)]$, contradicting Assumption 2.

Finally, it is obvious that the lagged values y_{it-s}, $i = 1, \ldots, v$, $i \neq j$, $s = 1, \ldots, n$, are predetermined. That the same is true of the contemporaneous regressors y_{it}, $i = 1, \ldots, j-1$, follows from the fact that the structure is recursive and we can orthogonalize the innovation process whilst maintaining both the recursive structure and the moving average row degrees. To verify this, let $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{D}\mathbf{C}'$ denote the Choleski decomposition of $\boldsymbol{\Sigma}$ where $\mathbf{C}$ is lower triangular with unit leading diagonal elements and $\mathbf{D} = \mathrm{diag}(d_1{}^2 \ldots, d_v^2)$. Then $\mathbf{w}_t = \mathbf{C}^{-1}\boldsymbol{\varepsilon}_t$ is a martingale difference innovation sequence with covariance matrix $\mathbf{D}$ and we can rewrite the ARMA system in (1.2) as

$$\mathbf{A}_0\mathbf{y}_t + \sum_{s=1}^{p} \mathbf{A}_s\mathbf{y}_{t-s} = \mathbf{L}_0\mathbf{w}_t + \sum_{s=1}^{p} \mathbf{L}_s\mathbf{w}_{t-s}. \tag{A.3}$$

where $\mathbf{L}_0 = \mathbf{M}_0\mathbf{C}$ is again lower triangular with unit leading diagonal elements, and because pre–multiplication by $\mathbf{M}_s$ reproduces zero-row structure, each $\mathbf{L}_s = \mathbf{M}_s\mathbf{C}$, $s = 1, \ldots, n$, has the same null rows as the corresponding $\mathbf{M}_s$. From the jth equation of (A.3) we now have

$$u_{jt} = w_{jt} + \sum_{i=1}^{j-1} l_{ji,0}w_{it} + \sum_{i=1}^{v}\sum_{s=1}^{n_j} l_{ji,s}w_{it-s}. \tag{A.4}$$

Thus, from (A.1) and (A.4) we see that in addition to w_{it}, the variable y_{it} depends at most on $w_{1t}, \ldots, w_{i-1t}$. Since by construction the elements of $\mathbf{w}_t$ are mutually uncorrelated, it follows that for $i = 1, \ldots, j-1$ we have $E[y_{it}w_{jt} \mid F_{t-1}] = 0$, as required. $\square$

Proof of Theorem 5.1:   The proof of Theorem 5.1 is based on the following lemmas.

Lemma A.1 *Suppose that $\mathbf{y}_t$ is an ARMA process satisfying Assumptions 1 and 2′ and assume that $(r(i), n_{r(i)}) = (r(i)^0, n_{r(i)}^0)$ for $i = j+1, \ldots, v$. Let $\widetilde{\mathbf{a}}^\dagger(z) = \sum_{s=0}^{n} \widetilde{\mathbf{a}}_s^\dagger z^s$ where the coefficients $\widetilde{\mathbf{a}}_s^\dagger$, $s = 0, \ldots, n$ belong to the solution set of the equation system*

$$\sum_{s=0}^{n} \widetilde{\mathbf{a}}_s^\dagger \widetilde{\boldsymbol{\Gamma}}_y(r+s) = \mathbf{0}, \; r = n+1, \ldots, 2n. \tag{A.5}$$

Set $\widetilde{\alpha}^\dagger(z) = \widetilde{\mathbf{a}}^\dagger(z)\mathbf{e}_j$ and $\widetilde{\boldsymbol{\beta}}^\dagger(z) = \widetilde{\mathbf{a}}^\dagger(z)(\mathbf{I} - \mathbf{e}_j\mathbf{e}_j')$. Given $\widetilde{\alpha}^\dagger(z)$ and $\widetilde{\boldsymbol{\beta}}^\dagger(z)$, let $\widetilde{\mu}^\dagger(z)$ be formed

from

$$\int_{-\pi}^{\pi} \sum_{s=0}^{n} \widetilde{\mu}_s^{\dagger} \frac{\widetilde{\mathbf{a}}^{\dagger}(\omega) \widetilde{\mathbf{S}}_y(\omega) \widetilde{\mathbf{a}}^{\dagger}(\omega)^*}{\widetilde{S}_v^{\dagger}(\omega)^2} \exp^{\imath \omega(s-r)} d\omega = 0, \ r = 1, \ldots, n \,, \tag{A.6}$$

where

$$\widetilde{S}_v^{\dagger}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-n}^{n} \widetilde{\mathbf{a}}^{\dagger}(\theta) \widetilde{\mathbf{S}}_y(\theta) \widetilde{\mathbf{a}}^{\dagger}(\theta)^* \exp^{\imath(\omega-\theta)r} d\theta \,. \tag{A.7}$$

Then, provided that for $i = j+1, \ldots, v$ the Kronecker invariant pairs satisfy $\{r(i)_T, n_{r(i)T}\} = \{r(i)^0, n_{r(i)}^0\}$, we have:

(i) *If $n < n_{r(j)0}$, $\widehat{\widetilde{\alpha}}(z) = \widetilde{\alpha}^{\dagger}(z) + O(Q_T)$, $\widehat{\widetilde{\beta}}(z) = \widetilde{\beta}^{\dagger}(z) + O(Q_T)$ and $\widehat{\widetilde{\mu}}(z) = \widetilde{\mu}^{\dagger}(z) + O(Q_T)$ uniformly in $|z| \leq 1$ where $Q_T = (loglogT/T)^{1/2}$;*

(ii) *If $n = n_{r(j)0}$, $\widehat{\widetilde{\alpha}}(z) = \widetilde{\alpha}^0(z) + O(Q_T)$, $\widehat{\widetilde{\beta}}(z) = \widetilde{\beta}^0(z) + O(Q_T)$ and $\widehat{\widetilde{\mu}}(z) = \widetilde{\mu}^0(z) + O(Q_T)$ uniformly in $|z| \leq 1$;*

(iii) *If $n > n_{r(j)0}$, $\widehat{\widetilde{\alpha}}(z) = \tilde{\phi}(z)\widetilde{\alpha}^0(z) + O(Q_T)$, $\widehat{\widetilde{\beta}}(z) = \tilde{\phi}(z)\widetilde{\beta}^0(z) + O(Q_T)$ and $\widehat{\widetilde{\mu}}(z) = \tilde{\phi}(z)\widetilde{\mu}^0(z) + O(Q_T)$ uniformly in $|z| \leq 1$ where $\tilde{\phi}(z) = 1 + \tilde{\phi}_1 z + \ldots + \tilde{\phi}_r z^r$, $r = n - n_{r(j)0}$, and $\tilde{\phi}(z) \neq 0$, $|z| \leq 1$.*

Proof of Lemma A.1: From Theorem 5.3.2 of Hannan and Deistler (1988) we know that $\|\widetilde{\mathbf{C}}_y(r) - \widetilde{\mathbf{\Gamma}}_y(r)\| = O(Q_T)$, $r = 0, \ldots, H_T \leq (\log T)^a$, $a < \infty$. From (4.1) and (A.5) we recognize that $\widehat{\widetilde{\mathbf{a}}}(z)$ and $\widetilde{\mathbf{a}}^{\dagger}(z)$ correspond to the solutions of systems of linear equations in which the coefficient matrices of the two systems differ by terms of order $O(Q_T)$. Moreover, the coefficient matrix of the system of equations is nonsingular by a direct application of Theorem 6.2.5 of Hannan and Deistler (1988). We can therefore conclude that $\widehat{\widetilde{\mathbf{a}}}_s - \widetilde{\mathbf{a}}_s^{\dagger} = O(Q_T)$, $s = 0, \ldots, n$. See also Theorem 6.2.4 of Hannan and Deistler (1988). Treating the polynomial operators as elements of the Hardy space $\mathcal{H}_2$, the space of functions analytic for $|z| < 1$ and square integrable on $|z| = 1$, now yields the result that

$$\|\widehat{\widetilde{\alpha}}(z) - \widetilde{\alpha}^{\dagger}(z)\|^2 = \|(\widehat{\widetilde{\mathbf{a}}}(z) - \widetilde{\mathbf{a}}^{\dagger}(z))\mathbf{e}_j\|^2 = \sum_{s=0}^{n}(\widehat{\widetilde{\alpha}}_s - \widetilde{\alpha}_s^{\dagger})^2 = O(Q_T^2)$$

and

$$\|\widehat{\widetilde{\beta}}(z) - \widetilde{\beta}^{\dagger}(z)\|^2 = \|(\widehat{\widetilde{\mathbf{a}}}^{\dagger}(z) - \widetilde{\mathbf{a}}^{\dagger}(z))(\mathbf{I} - \mathbf{e}_j\mathbf{e}_j')\|^2$$
$$= \sum_{i=1}^{j-1}(\widehat{\widetilde{\beta}}_i - \widetilde{\beta}_i^{\dagger})^2 + \sum_{\substack{i=1 \\ i \neq j}}^{v} \sum_{s=1}^{n}(\widehat{\widetilde{\beta}}_{i,s} - \widetilde{\beta}_{i,s}^{\dagger})^2 = O(Q_T^2) \,.$$

A parallel argument to that just employed will show that $\widehat{\widetilde{\mu}}(z) = \widetilde{\mu}^{\dagger}(z) + O(Q_T)$, $|z| \leq 1$, if it can be verified that the difference in the coefficient matrices of the two equation systems that define the operators $\widehat{\widetilde{\mu}}(z)$ and $\widetilde{\mu}^{\dagger}(z)$ are, likewise, $O(Q_T)$. To establish the latter, observe

that

$$
\begin{aligned}
\widetilde{C}_v(r) &= \sum_{s=0}^{n}\sum_{u=0}^{n} \widehat{\widetilde{\mathbf{a}}}_s \widetilde{\mathbf{C}}_y(r+s-u)\widehat{\widetilde{\mathbf{a}}}_u' \\
&= \sum_{s=0}^{n}\sum_{u=0}^{n} \left(\widetilde{\mathbf{a}}_s^{\dagger} + O(Q_T)\right)\left(\widetilde{\boldsymbol{\Gamma}}_y(r+s-u) + O(Q_T)\right)\left(\widetilde{\mathbf{a}}_u^{\dagger} + O(Q_T)\right)' \\
&= \sum_{s=0}^{n}\sum_{u=0}^{n} \widetilde{\mathbf{a}}_s^{\dagger}\widetilde{\boldsymbol{\Gamma}}_y(r+s-u)\widetilde{\mathbf{a}}_u^{\dagger\prime} + O(Q_T) \\
&= \widetilde{\gamma}_v^{\dagger}(r) + O(Q_T), \quad \text{say}, \tag{A.8}
\end{aligned}
$$

where $\widetilde{\gamma}_v^{\dagger}(r) = \int_{-\pi}^{\pi} \widetilde{\mathbf{a}}^{\dagger}(\omega)\widetilde{\mathbf{S}}_y(\omega)\widetilde{\mathbf{a}}^{\dagger}(\omega)^* \exp^{-\imath\omega r} d\omega$. From (A.8) it follows directly that

$$
\begin{aligned}
\widehat{\widetilde{S}}_v(\omega) &= \frac{1}{2\pi}\sum_{s=-n}^{n} \widetilde{C}_v(s)\exp(\imath\omega s) \\
&= \frac{1}{2\pi}\sum_{s=-n}^{n} \left(\widetilde{\gamma}_v^{\dagger}(s) + O(Q_T)\right)\exp(\imath\omega s) \\
&= \widetilde{S}_v^{\dagger}(\omega) + O(Q_T)
\end{aligned}
$$

uniformly in $\omega \in [-\pi, \pi]$.

Let $\widehat{\widetilde{\mathbf{H}}}(\omega) = \widehat{\widetilde{\mathbf{a}}}(\omega)/|\widehat{\widetilde{S}}_v(\omega)|$ and set $\widetilde{\mathbf{H}}^{\dagger}(\omega) = \widetilde{\mathbf{a}}^{\dagger}(\omega)/|\widetilde{S}_v^{\dagger}(\omega)|$. It is established below that $\widehat{\widetilde{\mathbf{H}}}(\omega)$ and $\widetilde{\mathbf{H}}^{\dagger}(\omega)$ belong to $\mathcal{L}_2$, with Fourier coefficients that decline at a geometric rate, and that $\int_{-\pi}^{\pi}\|\widehat{\widetilde{\mathbf{H}}}(\omega) - \widetilde{\mathbf{H}}^{\dagger}(\omega)\|^2 d\omega = O(Q_T^2)$. Now set

$$
\widetilde{\gamma i}_v^{\dagger}(r) = \int_{-\pi}^{\pi} \widetilde{\mathbf{H}}^{\dagger}(\omega)\widetilde{\mathbf{S}}_y(\omega)\widetilde{\mathbf{H}}^{\dagger}(\omega)^* \exp(-\imath\omega r)d\omega.
$$

Then, by definition,

$$
|\widetilde{C i}_v(r) - \widetilde{\gamma i}_v^{\dagger}(r)| = \left| \int_{-\pi}^{\pi}\left(\widehat{\widetilde{\mathbf{H}}}(\omega)\widetilde{\mathbf{I}}_y(\omega)\widehat{\widetilde{\mathbf{H}}}(\omega)^* - \widetilde{\mathbf{H}}^{\dagger}(\omega)\widetilde{\mathbf{S}}_y(\omega)\widetilde{\mathbf{H}}^{\dagger}(\omega)^*\right)\exp(-\imath\omega r)d\omega\right|, \quad \text{(A.9)}
$$

and suppressing the argument ω for convenience we have

$$
\begin{aligned}
\widehat{\widetilde{\mathbf{H}}}\widetilde{\mathbf{I}}_y\widehat{\widetilde{\mathbf{H}}}^* - \widetilde{\mathbf{H}}^{\dagger}\widetilde{\mathbf{S}}_y\widetilde{\mathbf{H}}^{\dagger*} =& (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})\widetilde{\mathbf{I}}_y(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* + (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})\widetilde{\mathbf{I}}_y\widetilde{\mathbf{H}}^{\dagger*} \\
& \widetilde{\mathbf{H}}^{\dagger}\widetilde{\mathbf{I}}_y(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* + \widetilde{\mathbf{H}}^{\dagger}(\widetilde{\mathbf{I}}_y - \widetilde{\mathbf{S}}_y)\widetilde{\mathbf{H}}^{\dagger*}. \tag{A.10}
\end{aligned}
$$

Substituting (A.10) into (A.9) we can now bound $|\widetilde{C i}_v(r) - \widetilde{\gamma i}_v^{\dagger}(r)|$ by the sum of four terms.

The first term is

$$
\left| \int_{-\pi}^{\pi}(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})\widetilde{\mathbf{I}}_y(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* \exp(-\imath\omega r)d\omega\right| \le \int_{-\pi}^{\pi}(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})\widetilde{\mathbf{I}}_y(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* d\omega. \tag{A.11}
$$

Applying the Cauchy-Schwartz inequality to the right hand side integrand in (A.11), recognizing that $\widetilde{\mathbf{I}}_y$ is Hermitian positive semi-definite, we find that

$$
\begin{aligned}
\int_{-\pi}^{\pi} (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger}) \widetilde{\mathbf{I}}_y (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* d\omega \;&=\; \frac{1}{2\pi T} \int_{-\pi}^{\pi} |(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger}) \widetilde{\mathbf{Z}}_y(\omega)|^2 d\omega \\
&\leq\; \int_{-\pi}^{\pi} \|(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})\|^2 \mathrm{tr}\{\widetilde{\mathbf{I}}_y\} d\omega \\
&\leq\; \sup_{\omega \in [-\pi, \pi]} \mathrm{tr}\{\widetilde{\mathbf{I}}_y\} \int_{-\pi}^{\pi} \|(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})\|^2 d\omega \\
&=\; O(\log T) O(Q_T^2),
\end{aligned}
$$

since $\limsup_{T \to \infty} [\sup_{\omega \in [-\pi, \pi]} \mathrm{tr}\{\widetilde{\mathbf{I}}_y(\omega)\} / \log T] \leq \sup_{\omega \in [-\pi, \pi]} 2\mathrm{tr}\{\widetilde{\mathbf{S}}_y(\omega)\}$ (Brillinger, 1975, Theorem 5.3.2.) Similarly, the second term, $\left| \int_{-\pi}^{\pi} (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger}) \widetilde{\mathbf{I}}_y \widetilde{\mathbf{H}}^{\dagger *} \exp(-\imath \omega r) d\omega \right|$, and the third term, $\left| \int_{-\pi}^{\pi} \widetilde{\mathbf{H}}^{\dagger} \widetilde{\mathbf{I}}_y (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* \exp(-\imath \omega r) d\omega \right|$, are both bounded by

$$
\begin{aligned}
\int_{-\pi}^{\pi} |\widetilde{\mathbf{H}}^{\dagger} \widetilde{\mathbf{I}}_y (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^*| \;&\leq\; \int_{-\pi}^{\pi} \left(\widetilde{\mathbf{H}}^{\dagger} \widetilde{\mathbf{I}}_y \widetilde{\mathbf{H}}^{\dagger *} \right)^{1/2} \left((\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger}) \widetilde{\mathbf{I}}_y (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* \right)^{1/2} d\omega \\
&\leq\; \int_{-\pi}^{\pi} \left(\widetilde{\mathbf{H}}^{\dagger} \widetilde{\mathbf{I}}_y \widetilde{\mathbf{H}}^{\dagger *} \right) d\omega \int_{-\pi}^{\pi} \left((\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger}) \widetilde{\mathbf{I}}_y (\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})^* \right) d\omega \\
&\leq\; (\sup_{\omega \in [-\pi, \pi]} \mathrm{tr}\{\widetilde{\mathbf{I}}_y\})^2 \int_{-\pi}^{\pi} \|\widetilde{\mathbf{H}}^{\dagger}\|^2 d\omega \int_{-\pi}^{\pi} \|(\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^{\dagger})\|^2 d\omega \\
&=\; (O(\log T))^2 O(Q_T^2).
\end{aligned}
$$

The fourth term is $\left| \int_{-\pi}^{\pi} \widetilde{\mathbf{H}}^{\dagger} (\widetilde{\mathbf{I}}_y - \widetilde{\mathbf{S}}_y) \widetilde{\mathbf{H}}^{\dagger *} \exp^{-\imath \omega r} d\omega \right|$, which equals

$$
\begin{aligned}
\Bigg| \sum_{u=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \widetilde{\mathbf{h}}_u^{\dagger} [\widetilde{\mathbf{C}}_y(r+u-s) &- \widetilde{\boldsymbol{\Gamma}}_y(r+u-s)] \widetilde{\mathbf{h}}_s^{\dagger \prime} \Bigg| \\
&\leq\; \sum_{u=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \left| \widetilde{\mathbf{h}}_u^{\dagger} [\widetilde{\mathbf{C}}_y(r+u-s) - \widetilde{\boldsymbol{\Gamma}}_y(r+u-s)] \widetilde{\mathbf{h}}_s^{\dagger \prime} \right| \\
&\leq\; \sum_{u=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \|\widetilde{\mathbf{h}}_u^{\dagger}\| \cdot \|\widetilde{\mathbf{C}}_y(r+u-s) - \widetilde{\boldsymbol{\Gamma}}_y(r+u-s)\| \cdot \|\widetilde{\mathbf{h}}_s^{\dagger}\|, \qquad \text{(A.12)}
\end{aligned}
$$

wherein we set $\widetilde{\mathbf{C}}_y(\tau) = 0$, $|\tau| \geq T$, and $\widetilde{\mathbf{h}}_u^{\dagger} = \int_{-\pi}^{\pi} \widetilde{\mathbf{H}}^{\dagger}(\omega) \exp^{-\imath \omega u} d\omega$, $u = 0, \pm 1, \ldots$. Since there exists a constant $\kappa > 0$ and a parameter λ, $0 < \lambda < 1$, such that $\|\widetilde{\mathbf{h}}_u^{\dagger}\| < \kappa \lambda^{|u|}$, the right hand side of equation (A.12) is less than or equal to

$$
\begin{aligned}
\kappa^2 \sum_{|u| < c \log T} \sum_{|s| < c \log T} &\|\widetilde{\mathbf{C}}_y(r+u-s) - \widetilde{\boldsymbol{\Gamma}}_y(r+u-s)\| \lambda^{|u|+|s|} \\
&+ \kappa^2 (\|\widetilde{\mathbf{C}}_y(0)\| + \|\widetilde{\boldsymbol{\Gamma}}_y(0)\|) \sum_{|u| \geq c \log T} \sum_{|s| \geq c \log T} \lambda^{|u|+|s|}. \qquad \text{(A.13)}
\end{aligned}
$$

By Theorem 5.3.2 of Hannan and Deistler (1988), the order of magnitude of the first term in this expression is $O(Q_T) 4\kappa^2 (1 - \lambda^{c \log T})^2 / (1 - \lambda)^2 = O(Q_T)$. The second term is bounded by a constant times $4\kappa^2 \lambda^{2c \log T} / (1 - \lambda)^2$, which is of order $O(T^{-1})$ for any $c \geq -1/(2 \log \lambda)$.

Hence we can conclude that the coefficient matrices of the two equation systems that define the operators $\widehat{\widetilde{\mu}}(z)$ and $\widetilde{\mu}^\dagger(z)$ differ by terms that are of order $O(Q_T)$ or smaller, as was required to be shown.

(i) The first statement of the lemma now follows without further ado.

(ii) When $n = n_{r(j)}^0$ it is readily verified that $\widetilde{\mathbf{a}}^\dagger(z) = \widetilde{\mathbf{a}}^0(z)$ provides the solution to (A.5). By definition $\widetilde{\mathbf{a}}^0(z) = \mathbf{e}_j' \widetilde{\mathbf{A}}^0(z)$, from which it follows that $\widetilde{\mathbf{a}}^0(\omega)\widetilde{\mathbf{K}}^0(\omega) = \mathbf{e}_j' \widetilde{\mathbf{M}}^0(\omega)$ where $\widetilde{\mathbf{a}}^0(\omega) = \widetilde{\mathbf{a}}^0(z)\big|_{z=e^{\imath\omega}}$ and, given that $\widetilde{\mathbf{S}}_y(\omega) = (2\pi)^{-1}\widetilde{\mathbf{K}}^0(\omega)\boldsymbol{\Sigma}^0\widetilde{\mathbf{K}}^0(\omega)^*$,

$$\widetilde{\mathbf{a}}^0(\omega)\widetilde{\mathbf{S}}_y(\omega) = (2\pi)^{-1}\mathbf{e}_j' \widetilde{\mathbf{M}}^0(\omega)\boldsymbol{\Sigma}^0\widetilde{\mathbf{K}}^0(\omega)^*\,. \tag{A.14}$$

From (A.14) we can conclude that $\int_{-\pi}^{\pi} \widetilde{\mathbf{a}}^0(\omega)\widetilde{\mathbf{S}}_y(\omega) \exp^{-\imath\omega u} d\omega = \mathbf{0}$ for $u > n$ and therefore

$$\int_{-\pi}^{\pi} \widetilde{\mathbf{a}}^0(\omega)\widetilde{\mathbf{S}}_y(\omega) \exp^{-\imath\omega r} d\omega = \sum_{s=0}^{n} \widetilde{\mathbf{a}}_s^0 \widetilde{\boldsymbol{\Gamma}}_y(r+s) = \mathbf{0}\,, \quad r = n+1, \ldots, 2n\,.$$

Similarly, the quadratic form

$$\begin{aligned}
\widetilde{\mathbf{a}}^0(\omega)\widetilde{\mathbf{S}}_y(\omega)\widetilde{\mathbf{a}}^0(\omega)^* &= (2\pi)^{-1}\mathbf{e}_j' \widetilde{\mathbf{M}}^0(\omega)\boldsymbol{\Sigma}^0\widetilde{\mathbf{M}}^0(\omega)^*\mathbf{e}_j \\
&= (2\pi)^{-1}(\sigma_\eta^0)^2|\widetilde{\mu}^0(\omega)|^2\,.
\end{aligned}$$

Substituting $\widetilde{\mathbf{a}}^\dagger(\omega) = \widetilde{\mathbf{a}}^0(\omega)$ into (A.7) we find that the equation system (A.6) corresponds to the Yule-Walker equations constructed from the (inverse) power spectrum $2\pi/(\sigma_\eta^0)^2|\widetilde{\mu}^0(\omega)|^2$, and hence that $\widetilde{\mu}^\dagger(z) = \widetilde{\mu}^0(z)$.

We are therefore lead to the conclusion that $\widehat{\widetilde{\alpha}}(z) = \alpha^o(z) + O(Q_T)$, $\widehat{\widetilde{\boldsymbol{\beta}}}(z) = \widetilde{\boldsymbol{\beta}}^o(z) + O(Q_T)$ and $\widehat{\widetilde{\mu}}(z) = \widetilde{\mu}^o(z) + O(Q_T)$, verifying the strong consistency claimed in Remark 1.

(iii) Now consider the case $n > n_{r(j)}^0$. Using the relationship in (A.14) it is straightforward to show that the solutions to Eq. (A.5) are characterized by operators of the form $\widetilde{\mathbf{a}}^\dagger(z) = \phi(z)\widetilde{\mathbf{a}}^0(z)$ where $\phi(z) = 1 + \phi_1 z + \ldots + \phi_r z^r$, $r = n - n_{r(j)}^0$. Moreover, since $\sum_{s=0}^{n} \widetilde{\mathbf{a}}_j^\dagger \widetilde{\boldsymbol{\Gamma}}_y(s-j) = \mathbf{0}$, $s \geq n$,

$$\left|\sum_{j=0}^{n} \widetilde{\mathbf{a}}_j^\dagger \widetilde{\mathbf{C}}_y(s-j)\right| \leq \sum_{j=0}^{n} \|\widetilde{\mathbf{a}}_j^\dagger\| \|\widetilde{\mathbf{C}}_y(s-j) - \widetilde{\boldsymbol{\Gamma}}_y(s-j)\| = O(Q_T)\sum_{j=0}^{n} \|\widetilde{\mathbf{a}}_j^\dagger\|.$$

Thus for T sufficiently large we can determine a measurable solution $\widehat{\widetilde{\mathbf{a}}}(z)$ to (4.1) such that, with arbitrarily large probability, $\widehat{\widetilde{\mathbf{a}}}(z) = \widehat{\phi}(z)\widetilde{\mathbf{a}}^0(z) + O(Q_T)$ where $\widehat{\phi}(z) \neq 0$, $|z| \leq 1$.

Substituting $\widetilde{\mathbf{a}}^\dagger(\omega) = \widehat{\phi}(\omega)\widetilde{\mathbf{a}}^0(\omega)$ into (A.7) we find that

$$\begin{aligned}
\widetilde{S}_v^\dagger(\omega) &= \frac{1}{2\pi}\sum_{r=-n}^{n}\int_{-\pi}^{\pi}|\widehat{\phi}(\theta)|^2\widetilde{\mathbf{a}}^0(\theta)\widetilde{\mathbf{S}}_y(\theta)\widetilde{\mathbf{a}}^0(\theta)^* \exp^{\imath(\omega-\theta)r} d\theta \\
&= \frac{(\sigma_\eta^0)^2}{2\pi}|\widehat{\phi}(\omega)|^2|\widetilde{\mu}^0(\omega)|^2
\end{aligned}$$

and hence, via (A.6), that $\widehat{\widetilde{\mu}}(z) = \widetilde{\mu}^\dagger(z) + O(Q_T)$ where the coefficients of $\widetilde{\mu}^\dagger(z)$ are derived from the Yule-Walker equations

$$\sum_{j=0}^{n} \widetilde{\mu}_j^\dagger \widetilde{\gamma} i_v^\dagger(r+j) = 0, \ r = 1, \ldots, n,$$

with

$$\widetilde{\gamma} i_v^\dagger(s) = \int_{-\pi}^{\pi} \frac{2\pi \exp^{-i\omega s}}{(\sigma_\eta^0)^2 |\widehat{\phi}(\omega)\widetilde{\mu}^0(\omega)|^2} \, d\omega \,.$$

This gives the desired conclusion.

To complete the proof it remains to be shown that $\widehat{\widetilde{\mathbf{H}}}, \widetilde{\mathbf{H}}^\dagger \in \mathcal{L}_2$, with Fourier coefficients that decline at a geometric rate, and that $\|\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^\dagger\|^2 = O(Q_T^2)$.

Consider $\widetilde{\mathbf{H}}^\dagger$. Given that $\widetilde{\mathbf{a}}^\dagger(z)$ is polynomial it is sufficient to show that $\widetilde{S}_v^\dagger(\omega)^{-2}$ is absolutely integrable to verify that $\widetilde{\mathbf{H}}^\dagger \in \mathcal{L}_2$. By Weierstrass's approximation theorem, for any $\epsilon > 0$, no matter how small, there exists a polynomial $p(z) = \sum_{j \geq 0} p_j z^j$, with $p(z) \neq 0$, $|z| \leq 1$, such that $|\widetilde{S}_v^\dagger(\omega)^2 - |p(\omega)|^2| \leq \epsilon$ uniformly in ω. Clearly, $p^{-1} \in \mathcal{L}_2$ and

$$
\begin{aligned}
\int_{-\pi}^{\pi} \frac{1}{\widetilde{S}_v^\dagger(\omega)^2 + \epsilon} d\omega &= \int_{-\pi}^{\pi} \frac{1}{|p(\omega)|^2} + \frac{1}{|p(\omega)|^2} \left\{ \frac{|p(\omega)|^2}{\widetilde{S}_v^\dagger(\omega)^2 + \epsilon} - 1 \right\} d\omega \\
&\leq \int_{-\pi}^{\pi} \frac{1}{|p(\omega)|^2} \left| 1 + \frac{\{|p(\omega)|^2 - \widetilde{S}_v^\dagger(\omega)^2 - \epsilon\}}{\widetilde{S}_v^\dagger(\omega)^2 + \epsilon} \right| d\omega \\
&\leq 3 \int_{-\pi}^{\pi} \frac{1}{|p(\omega)|^2} d\omega < \infty \,.
\end{aligned}
$$

Letting $\epsilon \to 0$ we can therefore conclude from Lebesgue's monotone convergence theorem that $\widetilde{S}_v^\dagger(\omega)^{-2}$ is absolutely integrable and hence that $\widetilde{\mathbf{H}}^\dagger \in \mathcal{L}_2$. Moreover, if $\widetilde{\mathbf{H}}^p(\omega) = \widetilde{\mathbf{a}}^\dagger(\omega)/|p(\omega)|$, then $\widetilde{\mathbf{H}}^p \in \mathcal{L}_2$ and $\widetilde{\mathbf{H}}^p$ can be expanded in a mean–square convergent Fourier series, $(2\pi)^{-1} \sum \widetilde{\mathbf{h}}_s^p \exp(\imath \omega s)$ say, where $\|\widetilde{\mathbf{h}}_s^p\| \to 0$ at a geometric rate as $|s| \to \infty$. Now,

$$
\begin{aligned}
\|\widetilde{\mathbf{H}}^\dagger - \widetilde{\mathbf{H}}^p\|^2 &= \int_{-\pi}^{\pi} \frac{\|\widetilde{\mathbf{a}}^\dagger\|^2}{\widetilde{S}_v^{\dagger 2}} \frac{\left| |p| - |\widetilde{S}_v^\dagger| \right|^2}{|p|^2} d\omega \\
&\leq \|\widetilde{\mathbf{H}}^\dagger\|^2 \sup_{[-\pi,\pi]} \frac{\epsilon^2}{|p|^4} \,,
\end{aligned}
\qquad (A.15)
$$

and $\|\widetilde{\mathbf{h}}_s^\dagger\| \leq \|\widetilde{\mathbf{h}}_s^p\| + \|\widetilde{\mathbf{h}}_s^\dagger - \widetilde{\mathbf{h}}_s^p\|$, implying that there exists a constant $\kappa > 0$ and a parameter λ, $0 < \lambda < 1$, such that $\|\widetilde{\mathbf{h}}_s^\dagger\| < \kappa \lambda^{|s|}$, since by Bessel's inequality and (A.15) $\sup_s \|\widetilde{\mathbf{h}}_s^\dagger - \widetilde{\mathbf{h}}_s^p\|^2 \leq \|\widetilde{\mathbf{H}}^\dagger - \widetilde{\mathbf{H}}^p\|^2 \to 0$ as $\epsilon \to 0$. A parallel proof, replacing $\widetilde{\mathbf{a}}^\dagger$ by $\widehat{\widetilde{\mathbf{a}}}$ and $\widetilde{S}_v^\dagger$ by $\widehat{\widetilde{S}}_v$, also shows that $\widehat{\widetilde{\mathbf{H}}} \in \mathcal{L}_2$, with Fourier coefficients that decline at a geometric rate.

Finally, $\|\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^\dagger\|^2 = O(Q_T^2)$ follows from the triangular inequality and the expansion

$$\widehat{\widetilde{\mathbf{H}}} - \widetilde{\mathbf{H}}^\dagger = (\widehat{\widetilde{\mathbf{a}}} - \widetilde{\mathbf{a}}^\dagger)\frac{1}{|\widehat{\widetilde{S}}_v|} + \widetilde{\mathbf{a}}^\dagger \left(\frac{|\widetilde{S}_v^\dagger| - |\widehat{\widetilde{S}}_v|}{|\widehat{\widetilde{S}}_v \widetilde{S}_v^\dagger|} \right)$$

because $\|\widehat{\widetilde{\mathbf{a}}}(\omega) - \widetilde{\mathbf{a}}^\dagger(\omega)\| = O(Q_T)$ and $\widehat{\widetilde{S}}_v(\omega) = \widetilde{S}_v^\dagger(\omega) + O(Q_T)$ uniformly in $\omega \in [-\pi, \pi]$. $\square$

Lemma A.2 *Suppose that $\mathbf{y}_t$ is an ARMA process satisfying Assumptions 1 and 2', and*

assume that for $j = q+1, \ldots, v$ the Kronecker invariant pairs $\{r(j)_T, n_{r(j)T}\} = \{r(j)^0, n_{r(j)}^0\}$. Then for all T sufficiently large

   (i) $\widehat{\widetilde{\sigma}}_\eta^2(n) > \widehat{\widetilde{\sigma}}_\eta^2(n+1)$ with probability one if $n < n_{r(q)}^0$, and

   (ii) if $n \geq n_{r(q)}^0$, $\widehat{\widetilde{\sigma}}_\eta^2(n) - \widehat{\widetilde{\sigma}}_\eta^2(n+1) = O(Q_T^2)$ almost surely.

Proof of Lemma A.2: As observed in Remark 2, Step 2 corresponds to a Gauss-Newton iteration designed to minimize

$$\frac{1}{T}\sum_t^T \widehat{\widetilde{\eta}}_t^2 = \int_{-\pi}^\pi \frac{\widehat{\widetilde{\mathbf{a}}}(\omega)\widetilde{\mathbf{I}}_y(\omega)\widehat{\widetilde{\mathbf{a}}}(\omega)^*}{|\widehat{\widetilde{\mu}}(\omega)|^2}d\omega + O((\log T/T^{1/2})). \tag{A.16}$$

Set

$$F_T(n) = \min_{\{\widetilde{\mathbf{a}}(z),\widetilde{\mu}(z)\}} \int_{-\pi}^\pi \frac{\widetilde{\mathbf{a}}(\omega)\widetilde{\mathbf{I}}_y(\omega)\widetilde{\mathbf{a}}(\omega)^*}{|\widetilde{\mu}(\omega)|^2}d\omega$$

where $\widetilde{\mathbf{a}}(z)$ and $\mu(z)$ are of degree n. Clearly, $F_T(n+1) \leq F_T(n)$. Now suppose that $F_T(n+1) = F_T(n)$. This implies that the function

$$\mathcal{F}(a,b) = \int_{-\pi}^\pi \frac{|1 - a\exp^{i\omega}|^2}{|1 - b\exp^{i\omega}|^2} \frac{|\widetilde{\mathbf{a}}(\omega)\widetilde{\mathbf{Z}}_y(\omega)|^2}{2\pi T|\widetilde{\mu}(\omega)|^2}d\omega$$

is minimised at $a = b$, $|b| < 1$. Following the method of argument used by Pötscher (1983, pp. 877–878) in the proof of his Theorem 5.2 we find that this leads to the conclusion that $\widetilde{\mathbf{a}}(z)\widetilde{\mathbf{Z}}_y(z)/\widetilde{\mu}(z) \equiv 1$ and the 'transfer function' $\widetilde{\mathbf{Z}}_y(z)$ is rational of degree n. But the order, or McMillan degree, of $\widetilde{\mathbf{Z}}_y(z) = \sum_{t=1}^T \widetilde{\mathbf{y}}_t z^t$ is T with probability one. Hence we can infer that $F_T(n+1) < F_T(n)$, *reductio ad absurdum*. Now set $0 < \delta_n < (F_T(n) - F_T(n+1))/4$. From (A.16), which follows from Theorems 4.5.2 and 5.3.2 of Brillinger (1975), it follows that when T is sufficiently large both the events $\widehat{\widetilde{\sigma}}_\eta^2(n) > F_T(n) - \delta_n$ and $\widehat{\widetilde{\sigma}}_\eta^2(n+1) < F_T(n+1) + \delta_n$ will occur with probability arbitrarily close to one. Thus $\widehat{\widetilde{\sigma}}_\eta^2(n) - \widehat{\widetilde{\sigma}}_\eta^2(n+1) > (F_T(n) - F_T(n+1)) - 2\delta_n > \frac{1}{2}(F_T(n) - F_T(n+1)) > 0$, confirming the result in (i) for $n < n_{r(q)}^0$.

We will now establish that $\widehat{\widetilde{\sigma}}_\eta^2(n) = T^{-1}\sum \widetilde{\eta}_t^2 + O(Q_T^2)$ whenever $n \geq n_{r(q)}^0$, from which the stated equality in (ii) follows directly. By definition,

$$\begin{aligned}
\widehat{\widetilde{\sigma}}_\eta^2(n) &= \frac{1}{T}\sum_t^T \widehat{\widetilde{\eta}}_t^2 + \sum_{j=0}^n \Delta\widehat{\widetilde{\mathbf{a}}}_j \frac{1}{T}\sum_{t=1}^T \widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\widetilde{\eta}}_t - \sum_{j=1}^n \Delta\widehat{\widetilde{\mu}}_j \frac{1}{T}\sum_{t=1}^T \widehat{\widetilde{\varphi}}_{t-j}\widehat{\widetilde{\eta}}_t \\
&= \frac{1}{T}\sum_t^T \widehat{\widetilde{\eta}}_t^2 + T^{-1}\sum_{t=1}^T (\Delta\widehat{\widetilde{\mathbf{a}}}(B)\widehat{\widetilde{\boldsymbol{\xi}}}_t - \Delta\widehat{\widetilde{\mu}}(B)\widehat{\widetilde{\varphi}}_t)\widehat{\widetilde{\eta}}_t, \tag{A.17}
\end{aligned}$$

the residual mean squared error from regressing $\widehat{\widetilde{\eta}}_t$ on the derivative processes $-\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}$, $j = 0, \ldots, n$, and $\widehat{\widetilde{\varphi}}_{t-j}$, $j = 1, \ldots, n$.

Consider first $T^{-1}\sum_t^T \widehat{\widetilde{\eta}}_t^2$. By Lemma A.1 $\widehat{\widetilde{\mathbf{a}}}(z) = \check{\phi}(z)\widetilde{\mathbf{a}}^0(z) + O(Q_T)$ and $\widehat{\widetilde{\mu}}(z) = \check{\phi}(z)\widetilde{\mu}^0(z) + O(Q_T)$ uniformly in $|z|$ where $\check{\phi}(z) \equiv 1$ when $n = n_{r(q)}^0$ and $\check{\phi}(z) \neq 0$, $|z| \leq 1$, when $n > n_{r(q)}^0$. Then

$$\frac{\widetilde{\mathbf{a}}^0(z)}{\widetilde{\mu}^0(z)} - \frac{\widehat{\widetilde{\mathbf{a}}}(z)}{\widehat{\widetilde{\mu}}(z)} = \sum_{j\geq 0} \psi_j^0 z^j$$

where the $\boldsymbol{\psi}_j^0$ are $O(Q_T)$ and decline at a geometric rate, so that $\|\boldsymbol{\psi}_j^0\| < O(Q_T)\lambda^j$, $j = 0, 1, 2, \ldots$, for some λ, $0 < \lambda < 1$. This implies that $|\tilde{\eta}_t - \widehat{\tilde{\eta}}_t| \leq \sum_{j\geq 0} \|\boldsymbol{\psi}_j^0\| \cdot \|\widetilde{\mathbf{y}}_{t-j}\| \leq O(Q_T) \sum_{j\geq 0} \lambda^j \|\widetilde{\mathbf{y}}_{t-j}\|$ and hence that $T^{-1} \sum (\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)^2 \leq O(Q_T^2)\mathrm{tr}\{\widetilde{\mathbf{C}}_y(0)\}/(1-\lambda)^2$. Furthermore,

$$T^{-1} \sum_{t=1}^T (\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)\tilde{\eta}_t = \sum_{j=0}^{a\log T} \boldsymbol{\psi}_j^0[T^{-1}\sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j}\tilde{\eta}_t] + R_T \tag{A.18}$$

where the remainder R_T is dominated by $O(Q_T)(\mathrm{tr}\{\widetilde{\mathbf{C}}_y(0)\}T^{-1}\sum \tilde{\eta}_t^2)^{1/2}\lambda^{a\log T}/(1-\lambda)$. From Proposition 3.1 we know that the variables in $\widetilde{\mathbf{y}}_{t-j}$, $j = 0, 1, \ldots, a\log T$ that are present in the mean cross products $T^{-1}\sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j}\tilde{\eta}_t$ occurring in (A.18) are predetermined relative to $\tilde{\eta}_t$. From Theorem 5.3.1 of Hannan and Deistler (1988) it follows that these mean cross products are $O(Q_T)$. This then leads to the conclusion that $T^{-1}\sum_{t=1}^T (\widehat{\tilde{\eta}}_t - \widehat{\tilde{\eta}}_{ct})\tilde{\eta}_t = O(Q_T^2)(1 - \lambda^{a\log T})/(1-\lambda) + O(Q_T)\lambda^{a\log T}/(1-\lambda)$. Taking $a > -1/\log\lambda$ gives the result that

$$\begin{aligned}T^{-1}\sum \widehat{\tilde{\eta}}_t^2 &= T^{-1}\sum \tilde{\eta}_t^2 + T^{-1}\sum(\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)^2 - 2T^{-1}\sum_{t=1}^T(\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)\tilde{\eta}_t \\ &= T^{-1}\sum \tilde{\eta}_t^2 + O(Q_T^2).\end{aligned} \tag{A.19}$$

Turning to the regression mean square, a similar argument to that just employed can be used to establish that when $n \geq n_{r(q)}^0$ the mean cross products of $\widehat{\tilde{\eta}}_t$ with the derivative processes $\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}$, $j = 0, \ldots, n$, and $\tilde{\varphi}_{t-j}$, $j = 1, \ldots, n$, are $O(Q_T)$. First,

$$T^{-1}\sum_{t=1}^T \widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\tilde{\eta}}_t = T^{-1}\sum_{s=0}^{a\log T} \widehat{\widetilde{k}}_s\{\sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j-s}\tilde{\eta}_t - \sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j-s}(\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)\} + R_T \tag{A.20}$$

where $\sum_{s\geq 0} \widehat{\widetilde{k}}_s z^s = 1/\widehat{\tilde{\mu}}(z)$ and $|\widehat{\widetilde{k}}_s| \leq \kappa\lambda^s$, $\kappa > 0$, $0 < \lambda < 1$. As above, Theorem 5.3.1 of Hannan and Deistler (1988) implies that $T^{-1}\sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j-s}\tilde{\eta}_t = O(Q_T)$, and from what has already been shown $\|T^{-1}\sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j-s}(\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)\| \leq (\mathrm{tr}\{\widetilde{\mathbf{C}}_y(0)\}T^{-1}\sum_{t=1}^T(\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)^2)^{1/2} = O(Q_T)$. The norm of the right hand side summations in (A.20) is therefore bounded by $O(Q_T)2\kappa(1 - \lambda^{a\log T})/(1-\lambda)$. The remainder $R_T = \sum_{s>a\log T} \widehat{\widetilde{k}}_s T^{-1}\sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j-s}\widehat{\tilde{\eta}}_t$ can be similarly bounded to give $\|R_T\| \leq \kappa(\mathrm{tr}\{\widetilde{\mathbf{C}}_y(0)\}T^{-1}\sum_{t=1}^T \widehat{\tilde{\eta}}_t^2)^{1/2}\lambda^{a\log T}/(1-\lambda)$. Taking $a > -1/\log\lambda$ implies that $\|R_T\| = O(T^{-1})$ and hence $T^{-1}\sum_{t=1}^T \widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\tilde{\eta}}_t = O(Q_T)$. Second,

$$T^{-1}\sum_{t=1}^T \widehat{\tilde{\varphi}}_{t-j}\widehat{\tilde{\eta}}_t = T^{-1}\sum_{s=0}^{a\log T} \widehat{\widetilde{\mathbf{k}}}_s\{\sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j-s}\tilde{\eta}_t - \sum_{t=1}^T \widetilde{\mathbf{y}}_{t-j-s}(\tilde{\eta}_t - \widehat{\tilde{\eta}}_t)\} + R_T \tag{A.21}$$

where $\sum_{s\geq 0} \widehat{\widetilde{\mathbf{k}}}_s z^s = \mathbf{a}(z)/\widehat{\tilde{\mu}}(z)^2$ and $\|\widehat{\widetilde{\mathbf{k}}}_s\| \leq \kappa\lambda^s$, $\kappa > 0$, $0 < \lambda < 1$. An argument that exactly parallels that used in the development immediately preceeding (A.21) leads to the conclusion that $T^{-1}\sum_{t=1}^T \widehat{\tilde{\varphi}}_{t-j}\widehat{\tilde{\eta}}_t = O(Q_T)$.

Given that the parameter adjustments $\Delta\widehat{\tilde{\mathbf{a}}}_j$, $j = 0, \ldots, n$, and $\Delta\widehat{\tilde{\mu}}_j$, $j = 1, \ldots, n$, can be expressed as weighted sums of $T^{-1}\sum \widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\tilde{\eta}}_t$, $j = 0, \ldots, n$, and $T^{-1}\sum \tilde{\varphi}_{t-j}\widehat{\tilde{\eta}}_t$, $j = 1, \ldots, n$, with weights that are almost surely $O(1)$, it follows that $\|\Delta\widehat{\tilde{\mathbf{a}}}(z)\|^2 = O(Q_T^2)$ and $\|\Delta\widehat{\tilde{\mu}}(z)\|^2 = O(Q_T^2)$.

Applying the Cauchy-Schwartz inequality we can therefore conclude that when $n \geq n^0_{r(q)}$ the regression mean square

$$T^{-1} \sum_{t=1}^{T} (\Delta \widehat{\widehat{\mathbf{a}}}(B)\widehat{\widehat{\boldsymbol{\xi}}}_t - \Delta\widehat{\mu}(B)\widehat{\varphi}_t)\widehat{\eta}_t = O(Q_T^2)\,. \tag{A.22}$$

Substituting (A.19) and (A.22) into (A.17) produces the desired result, namely that $\widehat{\widehat{\sigma}}^2_\eta(n) = T^{-1} \sum \tilde{\eta}_t^2 + O(Q_T^2)$ when $n \geq n^0_{r(q)}$. This completes the proof of Lemma (A.2). □

Proof of Theorem 5.1: Lemma (A.2) (i) implies that when $n < n^0_{r(q)}$

$$\liminf_{T\to\infty} \log\left[\widehat{\widehat{\sigma}}^2_\eta(n)/\widehat{\widetilde{\sigma}}^2_\eta(n+1)\right] > \log(1 + \widehat{\widehat{\rho}}(n)(1-\delta)) > 0$$

with probability one for any δ, $0 < \delta < 1$, where $\widehat{\widehat{\rho}}(n) = \widehat{\widehat{\sigma}}^2_\eta(n)/\widehat{\widetilde{\sigma}}^2_\eta(n+1) - 1 > 0$. The assumption that $p_T(d_q(n))/T \to 0$ a.s. as $T \to \infty$ therefore implies that $IC_T(n+1) < IC_T(n)$ a.s. for $n < n^0_{r(q)}$ because $p_T(d_q(n))/(T\log(1+\widehat{\widehat{\rho}}(1-\delta))) \to 0$ a.s.. Hence $\liminf_{T\to\infty} n_{r(q)T} \geq n^0_{r(q)}$. When $n \geq n^0_{r(q)}$, Lemma (A.2) (ii) implies

$$\log\left[\widehat{\widehat{\sigma}}^2_\eta(n)/\widehat{\widetilde{\sigma}}^2_\eta(n+1)\right] = \log(1 + O(Q_T^2)) = O(Q_T^2)\,.$$

If $\liminf_{T\to\infty} p_T(d_q(n))/L(T) > 0$, then since $d_q(n) < d_q(n+1)$

$$\frac{IC_T(n) - IC_T(n+1)}{L(T)} = \frac{TO(Q_T^2)}{L(T)} + \frac{p_T(d_q(n)) - p_T(d_q(n+1))}{L(T)} < 0\,,$$

implying that $IC_T(n + 1) > IC_T(n)$ a.s. for $n \geq n^0_{r(q)}$ and $\limsup_{T\to\infty} n_{r(q)T} \leq n^0_{r(q)}$. Theorem (5.1) follows on directly. □

Proof of Proposition 6.1: Since Proposition 3.1 depends on Assumption 2 and not the more restrictive Assumption 2′ it holds in its entirety for the $ARMA_E(\nu)$ representation of the levels $\mathbf{y}_t$ when $\mathbf{y}_t$ is an $ECARMA_E(\nu, \rho)$ cointegrated process. In order to couch Proposition 3.1 in terms of the EC representation in (6.1) we apply the Beveridge-Nelson decomposition before proceeding as in the development of (3.2) to give

$$\mathbf{b}(B)\triangle\mathbf{y}_t + \boldsymbol{\pi}'\mathbf{y}_{t-1} = \eta_t + \sum_{s=1}^{n} \mu_s\eta_{t-s}\,, \tag{A.23}$$

where $\mathbf{b}(z) = \mathbf{e}'_j\mathbf{B}(z)$ and $\boldsymbol{\pi}' = \mathbf{e}'_j\mathbf{CE}$. Now set $\triangle z_t = \mathbf{e}'_j\triangle\mathbf{y}_t$. Then given $\mathbf{T}_c$ we can manipulate (A.23) to produce expression (6.2) where $\alpha(z) = \mathbf{b}(z)\mathbf{e}_j$, $\boldsymbol{\beta}(z) = \mathbf{b}(z)(\mathbf{I} - \mathbf{e}_j\mathbf{e}'_j)$ and $\mathbf{c}' = \boldsymbol{\pi}'\mathbf{T}'_c(\mathbf{T}_c\mathbf{T}'_c)^{-1} = \mathbf{e}'_j\mathbf{F}$. Recall that in Proposition 3.1 the variables are assumed to have been ordered according to the Kronecker invariants, and note that $\mathbf{x}_{ct} = \mathbf{T}_c\mathbf{y}_t$ is invariant to the ordering of the variables in $\mathbf{y}_t$. □

Proof of Theorem 8.1: In what follows we will use the appendage of the subscript c to denote quantities calculated by applying Stage 2 of Algorithm–$ARMA_E(\nu, \varrho)$ to $\widetilde{\mathbf{z}}_t = [\triangle\widetilde{\mathbf{y}}'_t, \mathbf{x}'_{c(t-1)}]'$ rather than $\widetilde{\mathbf{w}}_t = [\triangle\widetilde{\mathbf{y}}'_t, \boldsymbol{\zeta}'_{(t-1)}]'$. Set $\mathbf{D} = \text{diag}(\mathbf{I}, \mathbf{R}_T)$. Then $\widetilde{\mathbf{w}}_t - \mathbf{D}\widetilde{\mathbf{z}}_t =$

$(\mathbf{0}, (\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)\mathbf{y}_{t-1})$ and

$$T^{-1}\sum_{t=1}^{T}\|\widetilde{\mathbf{w}}_t - \mathbf{D}\widetilde{\mathbf{z}}_t\|^2 = T^{-1}\mathrm{tr}[(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)\sum_{t=1}^{T}\mathbf{y}_{t-1}\mathbf{y}_{t-1}'(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)']$$
$$= O(Q_T^2) \qquad\qquad (A.24)$$

where the order of magnitude in (A.24) follows by Lemma 3.1 and Lemma A1 (i) of Poskitt (2000). Furthermore, $\widetilde{\mathbf{C}}_w(r) - \mathbf{D}\widetilde{\mathbf{C}}_z(r)\mathbf{D}'$ equals

$$\frac{1}{T}\left[\begin{array}{cc} \mathbf{0} & \sum_{t=r+1}^{T}\triangle\widetilde{\mathbf{y}}_t\mathbf{y}_{t-r-1}'(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)' \\ \sum_{t=r+1}^{T}(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)\mathbf{y}_{t-1}\triangle\widetilde{\mathbf{y}}_{t-r-1}' & \sum_{t=r+1}^{T}(\mathbf{B}_T\mathbf{y}_{t-1}\mathbf{y}_{t-r-1}'\mathbf{B}_T' - \mathbf{R}_T\mathbf{T}_c^0\mathbf{y}_{t-1}\mathbf{y}_{t-r-1}'\mathbf{T}_c^{0\prime}\mathbf{R}_T') \end{array}\right]$$

and using the Cauchy–Schwartz inequality in conjunction with (A.24) to bound the nonzero entries, having first rewritten the $(2,2)$ block as

$$T^{-1}\sum_{t=r+1}^{T}(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)\mathbf{y}_{t-1}\mathbf{y}_{t-r-1}'(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)' + T^{-1}\sum_{t=r+1}^{T}(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)\mathbf{y}_{t-1}\mathbf{y}_{t-r-1}'\mathbf{T}_c^{0\prime}\mathbf{R}_T'$$
$$+ T^{-1}\sum_{t=r+1}^{T}\mathbf{R}_T\mathbf{T}_c^0\mathbf{y}_{t-1}\mathbf{y}_{t-r-1}'(\mathbf{B}_T - \mathbf{R}_T\mathbf{T}_c^0)',$$

we find that

$$\widetilde{\mathbf{C}}_w(r) - \mathbf{D}\widetilde{\mathbf{C}}_z(r)\mathbf{D}' = \left[\begin{array}{cc} \mathbf{0} & O(Q_T) \\ O(Q_T) & O(Q_T) \end{array}\right] \qquad\qquad (A.25)$$

uniformly in r. Using equation (A.25) we can now establish the following lemma.

Lemma A.3 *Suppose that $\mathbf{y}_t$ is an $ECARMA_E(\nu, \rho)$ process satisfying Assumptions 1 and 2 and that $\varrho_T = \varrho^0$. Then $\|\widehat{\widetilde{\mathbf{d}}}(z) - \widehat{\widetilde{\mathbf{d}}}_c(z)\mathbf{D}^{-1}\|^2 = O(Q_T^2)$ and $\|\widehat{\widetilde{\mu}}(z) - \widehat{\widetilde{\mu}}_c(z)\|^2 = O(Q_T^2)$, $|z| \le 1$, where $\widehat{\widetilde{\mathbf{d}}}(z)$ and $\widehat{\widetilde{\mu}}(z)$, and $\widehat{\widetilde{\mathbf{d}}}_c(z)$ and $\widehat{\widetilde{\mu}}_c(z)$, denote the operators obtained from the coefficient values calculated by applying the first step of Stage 2 of Algorithm–$ARMA_E(\nu, \varrho)$ to $\widetilde{\mathbf{w}}_t$ and $\widetilde{\mathbf{z}}_t$ respectively.*

Proof of Lemma A.3: Replacing Theorem 5.3.2 of Hannan and Deistler (1988), as used in the proof of Lemma A.1, by (A.25), it follows from (7.1) that the coefficients of $\widehat{\widetilde{\mathbf{d}}}(z)$ and $\widehat{\widetilde{\mathbf{d}}}_c(z)\mathbf{D}^{-1}$ are derived from the solutions of nonsingular linear equation systems in which the terms differ by order $O(Q_T)$. We can therefore conclude that $\widehat{\widetilde{\mathbf{d}}}_s - \widehat{\widetilde{\mathbf{d}}}_{cs}\mathbf{D}^{-1} = O(Q_T)$, $s = 0, \ldots, n$, and hence that $\|\widehat{\widetilde{\mathbf{d}}}(z) - \widehat{\widetilde{\mathbf{d}}}_c(z)\mathbf{D}^{-1}\|^2 = O(Q_T^2)$, as required.

It now follows that

$$\begin{aligned}
\widetilde{C}_{cv}(r) &= \sum_{s=0}^{n}\sum_{u=0}^{n}\widehat{\widetilde{\mathbf{d}}}_{cs}\widetilde{\mathbf{C}}_z(r+s-u)\widehat{\widetilde{\mathbf{d}}}_{cu}' \\
&= \sum_{s=0}^{n}\sum_{u=0}^{n}\left(\widehat{\widetilde{\mathbf{d}}}_s + O(Q_T)\right)\left(\mathbf{D}\widetilde{\mathbf{C}}_z(r+s-u)\mathbf{D}'\right)\left(\widehat{\widetilde{\mathbf{d}}}_u + O(Q_T)\right)' \\
&= \sum_{s=0}^{n}\sum_{u=0}^{n}\widehat{\widetilde{\mathbf{d}}}_s\widetilde{\mathbf{C}}_w(r+s-u)\widehat{\widetilde{\mathbf{d}}}_u' + O(Q_T) \\
&= \widetilde{C}_v(r) + O(Q_T), \qquad\qquad (A.26)
\end{aligned}$$

and hence, by the same argument that follows (A.8), $\widehat{\widetilde{S}}_{cv}(\omega) = \widehat{\widetilde{S}}_v(\omega) + O(Q_T)$ uniformly in ω.

Let $\widehat{\widetilde{\mathbf{H}}}_c(\omega) = \widehat{\widetilde{\mathbf{d}}}_c(\omega)/|\widehat{\widetilde{S}}_{cv}(\omega)|$. By definition $\widetilde{C}i_{cv}(r) = \int_{-\pi}^{\pi} \widehat{\widetilde{\mathbf{H}}}_c(\omega)\widetilde{\mathbf{I}}_z(\omega)\widehat{\widetilde{\mathbf{H}}}_c(\omega)^* \exp(-\imath\omega r)d\omega$, and

$$|\widetilde{C}i_{cv}(r) - \widetilde{C}i_v(r)| = \left| \int_{-\pi}^{\pi} \left(\widehat{\widetilde{\mathbf{H}}}_c(\omega)\widetilde{\mathbf{I}}_z(\omega)\widehat{\widetilde{\mathbf{H}}}_c(\omega)^* - \widehat{\widetilde{\mathbf{H}}}(\omega)\widetilde{\mathbf{I}}_w(\omega)\widehat{\widetilde{\mathbf{H}}}(\omega)^* \right) \exp(-\imath\omega r)d\omega \right|. \quad \text{(A.27)}$$

where $\widehat{\widetilde{\mathbf{H}}}(\omega) = \widehat{\widetilde{\mathbf{d}}}(\omega)/|\widehat{\widetilde{S}}_v(\omega)|$. Suppressing the argument ω for convenience, we can substitute

$$\widehat{\widetilde{\mathbf{H}}}_c\widetilde{\mathbf{I}}_z\widehat{\widetilde{\mathbf{H}}}_c^* - \widehat{\widetilde{\mathbf{H}}}\widetilde{\mathbf{I}}_w\widehat{\widetilde{\mathbf{H}}}^* = \widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1}\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'\mathbf{D}'^{-1}\widehat{\widetilde{\mathbf{H}}}_c^* - \widehat{\widetilde{\mathbf{H}}}\widetilde{\mathbf{I}}_w\widehat{\widetilde{\mathbf{H}}}^*$$

$$= (\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^* + \widehat{\widetilde{\mathbf{H}}}(\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}' - \widetilde{\mathbf{I}}_w)\widehat{\widetilde{\mathbf{H}}}^*$$

$$+ \widehat{\widetilde{\mathbf{H}}}\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^* + (\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'\widehat{\widetilde{\mathbf{H}}}^* \quad \text{(A.28)}$$

into (A.27) and bound $|\widetilde{C}i_{cv}(r) - \widetilde{C}i_v(r)|$ by the sum of

$$I_1 = \left| \int_{-\pi}^{\pi} (\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^* \exp(-\imath\omega r)d\omega \right|,$$

$$I_2 = \left| \int_{-\pi}^{\pi} 2\Re\{\widehat{\widetilde{\mathbf{H}}}\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^*\} \exp(-\imath\omega r)d\omega \right|$$

and

$$I_3 = \left| \int_{-\pi}^{\pi} \widehat{\widetilde{\mathbf{H}}}(\widetilde{\mathbf{I}}_w - \mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}')\widehat{\widetilde{\mathbf{H}}}^* \exp(-\imath\omega r)d\omega \right|.$$

As in Lemma A.1, $\widehat{\widetilde{\mathbf{H}}}, \widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} \in \mathcal{L}_2$, with Fourier coefficients that decline at a geometric rate, and $\|\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}}\|^2 = O(Q_T^2)$. It follows that the first term

$$\begin{aligned}
I_1 &\leq \int_{-\pi}^{\pi} (\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^* d\omega \\
&\leq \int_{-\pi}^{\pi} \|\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}}\|^2 \text{tr}\{\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'\} d\omega \\
&\leq \sup_{\omega \in [-\pi, \pi]} \text{tr}\{\mathbf{D}\mathbf{I}_z\mathbf{D}'\}\|\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}}\|^2 \\
&= O(\log T)O(Q_T^2),
\end{aligned}$$

and the second

$$\begin{aligned}
I_2 &\leq 2\int_{-\pi}^{\pi} |\widehat{\widetilde{\mathbf{H}}}\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^*| d\omega \\
&\leq 2\int_{-\pi}^{\pi} \left(\widehat{\widetilde{\mathbf{H}}}\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'\widehat{\widetilde{\mathbf{H}}}^* \right)^{1/2} \left((\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^* \right)^{1/2} d\omega \\
&\leq 2\int_{-\pi}^{\pi} \left(\widehat{\widetilde{\mathbf{H}}}\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'\widehat{\widetilde{\mathbf{H}}}^* \right) d\omega \int_{-\pi}^{\pi} \left((\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'(\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}})^* \right) d\omega \\
&\leq 2(\sup_{\omega \in [-\pi, \pi]} \text{tr}\{\mathbf{D}\widetilde{\mathbf{I}}_z\mathbf{D}'\}^2 \|\widehat{\widetilde{\mathbf{H}}}\|^2 \|\widehat{\widetilde{\mathbf{H}}}_c\mathbf{D}^{-1} - \widehat{\widetilde{\mathbf{H}}}\|^2 \\
&= (O(\log T))^2 O(Q_T^2).
\end{aligned}$$

The third term

$$I_3 \leq \sum_{u=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \left| \widehat{\widetilde{\mathbf{h}}}_u [\widetilde{\mathbf{C}}_w(r+u-s) - \mathbf{D}\widetilde{\mathbf{C}}_z(r+u-s)\mathbf{D}'] \widehat{\widetilde{\mathbf{h}}}_s' \right| \tag{A.29}$$

wherein $\widetilde{\mathbf{C}}_w(\tau) = \widetilde{\mathbf{C}}_z(\tau) = \mathbf{0}$, $|\tau| \geq T$. Since $\|\widehat{\widetilde{\mathbf{h}}}_u\| < \kappa \lambda^{|u|}$ the right hand side of (A.29) can be bounded by

$$\kappa^2 \sum_{|u|<a\log T} \sum_{|s|<a\log T} \|\widetilde{\mathbf{C}}_w(r+u-s) - \mathbf{D}\widetilde{\mathbf{C}}_z(r+u-s)\mathbf{D}'\| \lambda^{|u|+|s|}$$
$$+ \kappa^2 (\|\widetilde{\mathbf{C}}_w(0)\| + \|\mathbf{D}\widetilde{\mathbf{C}}_z(0)\mathbf{D}\|) \sum_{|u|\geq a\log T} \sum_{|s|\geq a\log T} \lambda^{|u|+|s|}. \tag{A.30}$$

The first term in (A.30) is $O(Q_T)4\kappa^2(1-\lambda^{a\log T})^2/(1-\lambda)^2 = O(Q_T)$, by (A.25). The second term is dominated by a constant times $4\kappa^2\lambda^{2a\log T}/(1-\lambda)^2$, and taking $a \geq -1/(2\log\lambda)$ makes this term of order $O(T^{-1})$.

The difference in the coefficient matrices of the two systems of Yule–Walker equations that define the operators $\widehat{\widetilde{\mu}}(z)$ and $\widehat{\widetilde{\mu}}_c(z)$ have thus been shown to be $O(Q_T)$, and hence $\widehat{\widetilde{\mu}}(z) = \widehat{\widetilde{\mu}}_c(z) + O(Q_T)$, $|z| \leq 1$, giving the desired result. $\qquad \square$

Lemma A.4 *Suppose that $\mathbf{y}_t$ is an $ECARMA_E(\nu, \rho)$ process satisfying Assumptions 1 and 2, and assume that $\varrho_T = \varrho^0$ and that for $j = q+1, \ldots, v$ the Kronecker invariant pairs $\{r(j)_T, n_{r(j)T}\} = \{r(j)^0, n_{r(j)}^0\}$. If $\widehat{\widetilde{\sigma}}_\eta^2(n)$ and $\widehat{\widetilde{\sigma}}_{c\eta}^2(n)$ denote the mean squared error values obtained at the second step of Stage 2 of Algorithm–$ARMA_E(\nu, \varrho)$ when the algorithm is applied to $\widetilde{\mathbf{w}}_t$ and $\widetilde{\mathbf{z}}_t$ respectively, then for all T sufficiently large*

(i) $\widehat{\widetilde{\sigma}}_\eta^2(n) = \widehat{\widetilde{\sigma}}_{c\eta}^2(n) + O(Q_T)$ with probability one if $n < n_{r(q)}^0$, and

(ii) $\widehat{\widetilde{\sigma}}_\eta^2(n) = \widehat{\widetilde{\sigma}}_{c\eta}^2(n) + O(Q_T^2)$ almost surely if $n \geq n_{r(q)}^0$.

Proof of Lemma A.4: Let $\widehat{\widetilde{\eta}}_{ct}$ be defined as was $\widehat{\widetilde{\eta}}_t$ except that

$$\widehat{\widetilde{\eta}}_{ct} = \widehat{\widetilde{\mathbf{d}}}_c(B)\widetilde{\mathbf{z}}_t + (1 - \widehat{\widetilde{\mu}}_c(B))\widehat{\widetilde{\eta}}_{ct}.$$

Then the difference in the innovations estimates can be expressed as

$$\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct} = \frac{\widehat{\widetilde{\mathbf{d}}}(B)}{\widehat{\widetilde{\mu}}(B)} (\widetilde{\mathbf{w}}_t - \mathbf{D}\widetilde{\mathbf{z}}_t) + \left(\frac{\widehat{\widetilde{\mathbf{d}}}(B)}{\widehat{\widetilde{\mu}}(B)} - \frac{\widehat{\widetilde{\mathbf{d}}}_c(B)\mathbf{D}^{-1}}{\widehat{\widetilde{\mu}}_c(B)} \right) \mathbf{D}\widetilde{\mathbf{z}}_t.$$

Furthermore, Lemma A.3 implies that

$$\frac{\widehat{\widetilde{\mathbf{d}}}(z)}{\widehat{\widetilde{\mu}}(z)} - \frac{\widehat{\widetilde{\mathbf{d}}}_c(z)\mathbf{D}^{-1}}{\widehat{\widetilde{\mu}}_c(z)} = \sum_{j\geq 1} \widehat{\widetilde{\psi}}_{cj} z^j$$

where the $\widehat{\widetilde{\psi}}_{cj}$ are such that $\|\widehat{\widetilde{\psi}}_{cj}\| < O(Q_T)\lambda^j$, $j = 1, 2, \ldots$, for some λ, $0 < \lambda < 1$. Combining this result with (A.24) it therefore follows that $T^{-1}\sum(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})^2 = O(Q_T^2)$.

Similarly, the two derivative processes are such that

$$\widehat{\widetilde{\boldsymbol{\xi}}}_t - \mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{ct} = \frac{1}{\widehat{\widetilde{\mu}}(B)}\left(\widetilde{\mathbf{w}}_t - \mathbf{D}\widetilde{\mathbf{z}}_t\right) + \left(\frac{1}{\widehat{\widetilde{\mu}}(B)} - \frac{1}{\widehat{\widetilde{\mu}}_c(B)}\right)\mathbf{D}\widetilde{\mathbf{z}}_t$$

and

$$\widehat{\widetilde{\varphi}}_t - \widehat{\widetilde{\varphi}}_{ct} = \frac{1}{\widehat{\widetilde{\mu}}(B)}\left(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct}\right) + \left(\frac{1}{\widehat{\widetilde{\mu}}(B)} - \frac{1}{\widehat{\widetilde{\mu}}_c(B)}\right)\widehat{\widetilde{\eta}}_{ct},$$

from which we can conclude that $T^{-1}\sum\|\widehat{\widetilde{\boldsymbol{\xi}}}_t - \mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{ct}\|^2 = O(Q_T^2)$ and $T^{-1}\sum(\widehat{\widetilde{\varphi}}_t - \widehat{\widetilde{\varphi}}_{ct})^2 = O(Q_T^2)$.

It now follows that the mean squares and cross products of $\widehat{\widetilde{\eta}}_t$, $\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}$ and $\widehat{\widetilde{\varphi}}_{t-j}$, $j = 0, \ldots, n$, differ from those of $\widehat{\widetilde{\eta}}_{ct}$, $\mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}$ and $\widehat{\widetilde{\varphi}}_{c(t-j)}$, $j = 0, \ldots, n$, by terms of order $O(Q_T)$. For example,

$$\sum_{t=1}^{T}\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\widetilde{\eta}}_t - \sum_{t=1}^{T}\mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct} = \sum_{t=1}^{T}(\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j} - \mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)})(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct}) + \sum_{t=1}^{T}(\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j} - \mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)})\widehat{\widetilde{\eta}}_{ct}$$

$$+ \sum_{t=1}^{T}\mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct}),$$

from which the conclusion that $T^{-1}\sum\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\widetilde{\eta}}_t = T^{-1}\sum\mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct} + O(Q_T)$ can be deduced via application(s) of the Cauchy–Schwartz inequality.

Given that the difference in the various mean squares and cross products is $O(Q_T)$, the difference in the normal equations that define the operators $\Delta\widehat{\widetilde{\mathbf{d}}}(z)$ and $\Delta\widehat{\widetilde{\mu}}(z)$, and $\Delta\widehat{\widetilde{\mathbf{d}}}_c(z)\mathbf{D}^{-1}$ and $\Delta\widehat{\widetilde{\mu}}_c(z)$, are likewise $O(Q_T)$, and hence $\|\Delta\widehat{\widetilde{\mathbf{d}}}(z) - \Delta\widehat{\widetilde{\mathbf{d}}}_c(z)\mathbf{D}^{-1}\|^2 = O(Q_T^2)$ and $\|\Delta\widehat{\widetilde{\mu}}(z) - \Delta\widehat{\widetilde{\mu}}_c(z)\|^2 = O(Q_T^2)$, $|z| \leq 1$.

Now, by the Chain–Rule,

$$\frac{\partial\mathbf{d}_{cj}\mathbf{D}^{-1}}{\partial\mathbf{d}'_{cj}}\frac{\partial\sum_t^T\widehat{\widetilde{\eta}}_{ct}^2}{\partial(\mathbf{d}_{cj}\mathbf{D}^{-1})'} = \frac{\partial\sum_t^T\widehat{\widetilde{\eta}}_{ct}^2}{\partial\mathbf{d}'_{cj}} = 2\sum_{t=1}^{T}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct},$$

and

$$\frac{\partial\sum_t^T\widehat{\widetilde{\eta}}_{ct}^2}{\partial\mu_{cj}} = 2\sum_{t=1}^{T}\widehat{\widetilde{\varphi}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct}.$$

Expanding the difference $\widehat{\widetilde{\sigma}}^2_\eta(n) - \widehat{\widetilde{\sigma}}^2_{c\eta}(n)$ in terms of the differences in the coefficient adjustments and the associated derivative processes gives

$$
\frac{1}{T}\sum_t^T(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})^2 - \frac{2}{T}\sum_t^T(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})\widehat{\widetilde{\eta}}_{ct} + \sum_{j=0}^{n-1}(\Delta\widehat{\widetilde{\mathbf{d}}}_j - \Delta\widehat{\widetilde{\mathbf{d}}}_{cj}\mathbf{D}^{-1})\frac{1}{T}\sum_{t=1}^T(\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\widetilde{\eta}}_t - \mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct})
$$

$$
+ \sum_{j=0}^{n-1}(\Delta\widehat{\widetilde{\mu}}_j - \Delta\widehat{\widetilde{\mu}}_{cj})\frac{1}{T}\sum_{t=1}^T(\widehat{\widetilde{\varphi}}_{t-j}\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\varphi}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct}) + \sum_{j=0}^{n-1}(\Delta\widehat{\widetilde{\mathbf{d}}}_j - \Delta\widehat{\widetilde{\mathbf{d}}}_{cj}\mathbf{D}^{-1})\frac{1}{T}\sum_{t=1}^T\mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct}
$$

$$
+ \sum_{j=0}^{n-1}(\Delta\widehat{\widetilde{\mu}}_j - \Delta\widehat{\widetilde{\mu}}_{cj})\frac{1}{T}\sum_{t=1}^T\widehat{\widetilde{\varphi}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct} + \sum_{j=0}^{n-1}\Delta\widehat{\widetilde{\mathbf{d}}}_{cj}\mathbf{D}^{-1}\frac{1}{T}\sum_{t=1}^T(\widehat{\widetilde{\boldsymbol{\xi}}}_{t-j}\widehat{\widetilde{\eta}}_t - \mathbf{D}\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct})
$$

$$
+ \sum_{j=0}^{n-1}\Delta\widehat{\widetilde{\mu}}_{cj}\frac{1}{T}\sum_{t=1}^T(\widehat{\widetilde{\varphi}}_{t-j}\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\varphi}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct}), \tag{A.31}
$$

Recognizing that the previous derivations and almost sure bounds are applicable for all values of $n \geq n^0_{r(q+1)}$, we can now deduce, after application of the Cauchy–Schwartz inequality, that all the terms in (A.31) are of order $O(Q_T)$, or smaller, and Part (i) of Lemma A.4 follows on directly.

When $n \geq n^0_{r(q)}$ we can tighten the bounds on the terms in (A.31) by noting from Lemma A.5 below that $\widehat{\widetilde{\mathbf{d}}}_c(z) = \tilde{\phi}(z)\tilde{\mathbf{d}}^0(z) + O(Q_T)$ and $\widehat{\widetilde{\mu}}_c(z) = \tilde{\phi}(z)\tilde{\mu}^0(z) + O(Q_T)$ uniformly in $|z|$ where $\tilde{\phi}(z) \equiv 1$ when $n = n^0_{r(q)}$ and $\tilde{\phi}(z) \neq 0$, $|z| \leq 1$, when $n > n^0_{r(q)}$. Then

$$
\tilde{\eta}_t - \widehat{\widetilde{\eta}}_{ct} = \left(\frac{\tilde{\mathbf{d}}^0(B)}{\tilde{\mu}^0(B)} - \frac{\widehat{\widetilde{\mathbf{d}}}_c(B)}{\widehat{\widetilde{\mu}}_c(B)}\right)\widetilde{\mathbf{z}}_t = \sum_{j \geq 0}\boldsymbol{\psi}^0_{cj}\widetilde{\mathbf{z}}_{t-j}
$$

where $\|\boldsymbol{\psi}^0_{cj}\| < O(Q_T)\lambda^j$, $j = 1, 2, \ldots$, for some λ, $0 < \lambda < 1$. This implies that the mean squared differences $T^{-1}\sum(\tilde{\eta}_t - \widehat{\widetilde{\eta}}_{ct})^2 \leq O(Q_T^2)\mathrm{tr}\{\widetilde{\mathbf{C}}_z(0)\}(1-\lambda)^{-2}$. Similarly,

$$
T^{-1}\sum_{t=1}^T(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})\tilde{\eta}_t = \sum_{j=0}^{a\log T}\widehat{\widetilde{\psi}}_{cj}[T^{-1}\sum_{t=1}^T\widetilde{\mathbf{z}}_{t-j}\tilde{\eta}_t] + R_T \tag{A.32}
$$

where $|R_T| \leq O(Q_T)(\mathrm{tr}\{\widetilde{\mathbf{C}}_z(0)\}T^{-1}\sum_{t=1}^T\tilde{\eta}_t^2)^{1/2}\lambda^{a\log T}/(1-\lambda)$. From Proposition 6.1 we know that the variables in $\widetilde{\mathbf{z}}_{t-j}$, $j = 0, 1, \ldots, a\log T$ that occur in the mean cross products $T^{-1}\sum_{t=1}^T\widetilde{\mathbf{z}}_{t-j}\tilde{\eta}_t$ appearing in (A.32) are predetermined relative to $\tilde{\eta}_t$. From Theorem 5.3.1 of Hannan and Deistler (1988) it follows that these mean cross products are $O(Q_T)$. Taking $a > -1/\log\lambda$ implies that $|R_T| = O(Q_TT^{-1})$ and given the bounds on the mean cross products we can infer that $T^{-1}\sum_{t=1}^T(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})\tilde{\eta}_t = O(Q_T^2)(1 - \lambda^{a\log T})/(1-\lambda) + O(Q_TT^{-1})$. This then leads to the conclusion that $\frac{1}{T}\sum_t^T(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})\widehat{\widetilde{\eta}}_{ct} = \frac{1}{T}\sum_t^T(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})(\widehat{\widetilde{\eta}}_{ct} - \tilde{\eta}_t) + \frac{1}{T}\sum_{t=1}^T(\widehat{\widetilde{\eta}}_t - \widehat{\widetilde{\eta}}_{ct})\tilde{\eta}_t = O(Q_T^2)$.

Furthermore, in a manner analogous to (A.20) and (A.21), the mean cross products $T^{-1}\sum_{t=1}^T\widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct}$ and $T^{-1}\sum_{t=1}^T\tilde{\varphi}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct}$ can be expanded as weighted sums of $T^{-1}\sum\widetilde{\mathbf{z}}_{t-j-k}\tilde{\eta}_t$ and $T^{-1}\sum\widetilde{\mathbf{z}}_{t-j-k}[\tilde{\eta}_t - \widehat{\widetilde{\eta}}_{ct}]$, for $k = 0, \ldots, a\log T$, with weights that decline geometrically, plus a remainder that is $O(1)\lambda^{a\log T}/(1-\lambda)$. Using a repetition of the previous logic and employing arguments equivalent to those surrounding (A.20) and (A.21), it can be shown that

when $n \geq n_{r(q)}^0$ the first two components contribute terms that are $O(Q_T)$ and the remainder is $O(T^{-1})$. Hence $T^{-1} \sum_{t=1}^{T} \widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)} \widehat{\widetilde{\eta}}_{ct}$, $j = 0, \ldots, n$ and $T^{-1} \sum_{t=1}^{T} \tilde{\varphi}_{c(t-j)} \widehat{\widetilde{\eta}}_{ct}$, $j = 1, \ldots, n$, are $O(Q_T)$, from which it follows that $\Delta \widehat{\widetilde{\mathbf{d}}}_{cj}$, $j = 0, \ldots, n$, and $\Delta \widehat{\widetilde{\mu}}_{cj}$, $j = 1, \ldots, n$, are likewise $O(Q_T)$.

Applying these bounds to the terms in (A.31) we now find that when $n \geq n_{r(q)}^0$ (A.31) is $O(Q_T^2)$, which completes the proof of Part (ii) of Lemma A.4. □

Lemma A.5 *Suppose that* $\mathbf{y}_t$ *is an* $ECARMA_E(\nu, \rho)$ *process satisfying Assumptions 1 and 2. Assume also that* $\varrho_T = \varrho^0$ *and that for* $j = q+1, \ldots, v$ *the Kronecker invariant pairs* $\{r(j)_T, n_{r(j)T}\} = \{r(j)^0, n_{r(j)}^0\}$. *Let* $\widetilde{\mathbf{d}}_s^\dagger$, $s = 0, \ldots, n-1$, *belong to the solution set of the equation system*

$$\sum_{s=0}^{n} \widetilde{\mathbf{d}}_s^\dagger \widetilde{\boldsymbol{\Gamma}}_z(r+s) = \mathbf{0}, \ r = n, \ldots, 2(n-1), \tag{A.33}$$

set $\widetilde{\mathbf{d}}^\dagger(z) = \sum_{s=0}^{n-1} \widetilde{\mathbf{d}}_s^\dagger z^s$, *and let* $\widetilde{\mu}^\dagger(z)$ *be formed from*

$$\int_{-\pi}^{\pi} \sum_{s=0}^{n} \widetilde{\mu}_s^\dagger \frac{\widetilde{\mathbf{d}}^\dagger(\omega) \widetilde{\mathbf{S}}_z(\omega) \widetilde{\mathbf{d}}^\dagger(\omega)^*}{\widetilde{S}_v^\dagger(\omega)^2} \exp^{\imath \omega(s-r)} d\omega = 0, \ r = 1, \ldots, n, \tag{A.34}$$

where

$$\widetilde{S}_v^\dagger(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{r=-n}^{n} \widetilde{\mathbf{d}}^\dagger(\theta) \widetilde{\mathbf{S}}_z(\theta) \widetilde{\mathbf{d}}^\dagger(\theta)^* \exp^{\imath(\omega-\theta)r} d\theta. \tag{A.35}$$

Then uniformly in $|z| \leq 1$;

(i) $\widehat{\widetilde{\mathbf{d}}}_c(z) = \widetilde{\mathbf{d}}^\dagger(z) + O(Q_T)$ *and* $\widehat{\widetilde{\mu}}_c(z) = \widetilde{\mu}^\dagger(z) + O(Q_T)$ *if* $n < n_{r(q)}^0$,

(ii) $\widehat{\widetilde{\mathbf{d}}}_c(z) = \widetilde{\mathbf{d}}^0(z) + O(Q_T)$ *and* $\widehat{\widetilde{\mu}}_c(z) = \widetilde{\mu}^0(z) + O(Q_T)$ *if* $n = n_{r(q)}^0$, *and*

(iii) *if* $n > n_{r(q)}^0$, $\widehat{\widetilde{\mathbf{d}}}_c(z) = \tilde{\phi}(z)\widetilde{\mathbf{d}}^0(z) + O(Q_T)$ *and* $\widehat{\widetilde{\mu}}_c(z) = \tilde{\phi}(z)\widetilde{\mu}^0(z) + O(Q_T)$ *where* $\tilde{\phi}(z) = 1 + \tilde{\phi}_1 z + \ldots + \tilde{\phi}_r z^r$, $r = n - n_{r(j)0}$, *and* $\tilde{\phi}(z) \neq 0$, $|z| \leq 1$.

Proof of Lemma A.5: Upon making the translations $\widetilde{\mathbf{y}}_t \mapsto \widetilde{\mathbf{z}}_t$, $\widetilde{\mathbf{a}}(z) \mapsto \widetilde{\mathbf{d}}(z)$, $\widetilde{\mu}(z) \mapsto \widetilde{\mu}(z)$, $\widehat{\widetilde{\mathbf{a}}}(z) \mapsto \widehat{\widetilde{\mathbf{d}}}_c(z)$ and $\widehat{\widetilde{\mu}}(z) \mapsto \widehat{\widetilde{\mu}}_c(z)$, it is clear that Lemma A.5 provides a direct parallel to Lemma A.1, and having made the translations the proof of Lemma A.5 parallels that of Lemma A.1 in a similar manner. The detailed steps of the argument, which are akin to those already employed in the proof of Lemma A.3, are omitted. □

Lemma A.6 *Suppose that* $\mathbf{y}_t$ *is an* $ECARMA_E(\nu, \rho)$ *process satisfying Assumptions 1 and 2. Assume also that* $\varrho_T = \varrho^0$ *and that for* $j = q+1, \ldots, v$ *the Kronecker invariant pairs* $\{r(j)_T, n_{r(j)T}\} = \{r(j)^0, n_{r(j)}^0\}$. *Then for all* T *sufficiently large*

(i) $\widehat{\widetilde{\sigma}}_{c\eta}^2(n) > \widehat{\widetilde{\sigma}}_{c\eta}^2(n+1)$ *with probability one if* $n < n_{r(q)}^0$, *and*

(ii) *if* $n \geq n_{r(q)}^0$, $\widehat{\widetilde{\sigma}}_{c\eta}^2(n) = \widehat{\widetilde{\sigma}}_{c\eta}^2(n+1) + O(Q_T^2)$ *almost surely.*

Proof of Lemma A.6: Lemma A.6 parallels Lemma A.2 in the same way that Lemma A.5 parallels Lemma A.1, and on making the translations $\widetilde{\mathbf{y}}_t \mapsto \widetilde{\mathbf{z}}_t$, $\widetilde{\mathbf{a}}(z) \mapsto \widetilde{\mathbf{d}}(z)$, $\widetilde{\mu}(z) \mapsto \widetilde{\mu}(z)$,

$\widehat{\widetilde{\mathbf{a}}}(z) \mapsto \widehat{\widetilde{\mathbf{d}}}_c(z)$ and $\widehat{\widetilde{\mu}}(z) \mapsto \widehat{\widetilde{\mu}}_c(z)$, the proof of Lemma A.6 proceeds as in the proof of Lemma A.2.

Recognizing that without exception $\mathbf{x}_{c(t-1)}$ enters each candidate specification through the error correction term, we find that the incorporation of the cointegrating relations has no bearing on the proof of (i). This is so because

$$F_T(n|\varrho) = \min_{\{\widetilde{\mathbf{d}}(z),\widetilde{\mu}(z)\}} \int_{-\pi}^{\pi} \frac{\widetilde{\mathbf{d}}(\omega)\widetilde{\mathbf{I}}_z(\omega)\widetilde{\mathbf{d}}(\omega)^*}{|\widetilde{\mu}(\omega)|^2} d\omega$$

depends on n but does not depend on ϱ, which is held constant. The proof of Lemma A.6 (i) can therefore proceed along the same lines as the proof of Lemma 5.2 (i).

Similarly, the proof of Lemma A.6 (ii) follows that of Lemma A.2 (ii). Indeed, a substantial part of the proof of (ii) has already been given in the proof of Lemma A.3, where it was shown that $T^{-1}\sum(\tilde{\eta}_t - \widehat{\widetilde{\eta}}_{ct})^2 = O(Q_T^2)$, that $T^{-1}\sum_{t=1}^{T} \widehat{\widetilde{\boldsymbol{\xi}}}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct} = O(Q_T)$, $j = 0,\ldots,n$, and $T^{-1}\sum_{t=1}^{T} \tilde{\varphi}_{c(t-j)}\widehat{\widetilde{\eta}}_{ct} = O(Q_T)$, $j = 1,\ldots,n$, and that $\|\Delta\widehat{\widetilde{\mathbf{d}}}_c(z)\|^2 = O(Q_T^2)$ and $\|\Delta\widehat{\widetilde{\mu}}_c(z)\|^2 = O(Q_T^2)$ when $n \geq n_{r(q)}^0$. To complete the proof it is only necessary to add the expansion $T^{-1}\sum_{t=1}^{T}(\tilde{\eta}_t - \widehat{\widetilde{\eta}}_{ct})\tilde{\eta}_t = \sum_{j=0}^{a\log T} \boldsymbol{\psi}_{cj}^0[T^{-1}\sum_{t=1}^{T} \mathbf{z}_{t-j}\tilde{\eta}_t] + R_T$, which following the argument after (A.32) can be shown to be $O(Q_T^2)$. Hence we can conclude that $T^{-1}\sum_{t=1}^{T} \widehat{\widetilde{\eta}}_{ct}^2 = T^{-1}\sum_{t=1}^{T} \tilde{\eta}_t^2 + O(Q_T^2)$, from which the statement in Lemma A.6 (ii) follows. $\square$

Proof of Theorem 8.1: By Lemma A.4 (i), when $n < n_{r(q)}^0$

$$\log\left[\widehat{\widetilde{\sigma}}_\eta^2(n)/\widehat{\widetilde{\sigma}}_\eta^2(n+1)\right] = \log\left[\widehat{\widetilde{\sigma}}_{c\eta}^2(n)/\widehat{\widetilde{\sigma}}_{c\eta}^2(n+1) + O(Q_T)\right],$$

which implies via Lemma (A.6) (i) that

$$\liminf_{T\to\infty}\log\left[\widehat{\widetilde{\sigma}}_\eta^2(n)/\widehat{\widetilde{\sigma}}_\eta^2(n+1)\right] > \log(1 + \widehat{\widetilde{\rho}}_c(n)(1-\delta) + O(Q_T)) > 0$$

with probability one for any δ, $0 < \delta < 1$, where $\widehat{\widetilde{\rho}}_c(n) = \widehat{\widetilde{\sigma}}_{c\eta}^2(n)/\widehat{\widetilde{\sigma}}_{c\eta}^2(n+1) - 1 > 0$. We can therefore conclude that $IC_T(n+1) < IC_T(n)$ a.s. when $n < n_{r(q)}^0$ because the assumption $p_T(d_q(n))/T \to 0$ almost surely as $T \to \infty$ implies that $p_T(d_q(n))/(T\log(1 + \widehat{\widetilde{\rho}}(1-\delta) + O(Q_T))) \to 0$ a.s.. Hence $\liminf_{T\to\infty} n_{r(q)T} \geq n_{r(q)}^0$. When $n \geq n_{r(q)}^0$, Lemma A.4 (ii) and Lemma A.6 (ii) imply that

$$\begin{aligned}
\log\left[\widehat{\widetilde{\sigma}}_\eta^2(n)/\widehat{\widetilde{\sigma}}_\eta^2(n+1)\right] &= \log\left[\widehat{\widetilde{\sigma}}_{c\eta}^2(n)/\widehat{\widetilde{\sigma}}_{c\eta}^2(n+1) + O(Q_T^2)\right] \\
&= \log(1 + O(Q_T^2)) \\
&= O(Q_T^2).
\end{aligned}$$

If $\liminf_{T\to\infty} p_T(d_q(n))/L(T) > 0$ then since $d_q(n) < d_q(n+1)$

$$\frac{IC_T(n) - IC_T(n+1)}{L(T)} = \frac{TO(Q_T^2)}{L(T)} + \frac{p_T(d_q(n)) - p_T(d_q(n+1))}{L(T)} < 0,$$

implying that $IC_T(n+1) > IC_T(n)$. Theorem (8.1) now follows directly. $\square$